JPL D-7669, Part 1

# Planetary Data System
# Data Preparation Workbook

February 17, 1995
Version 3.1

PDS
Planetary Data System
National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

# Table of Contents

# Chapter 1

# Introduction

## 1.1 Purpose

This PDS Data Preparation Workbook (DPW) serves as a guide for the organization and preparation of data sets intended for submission to the Planetary Data System (PDS). This document describes procedures to be used for preparing a data set that meets PDS archive standards. It is intended to be used in conjunction with the PDS Standards Reference and the Planetary Science Data Dictionary, which contain specific PDS data preparation standards. The procedures provided in this volume reference the standards provided in the PDS Standards Reference. This document also describes software tools available from PDS that support the archive preparation process.

## 1.2 Audience

This workbook is intended for scientists and engineers in the planetary science community who are in the process of, or are considering, submitting restored or new mission data sets to PDS. It is designed primarily for data producers that are new to PDS. Persons already familiar with PDS may find this workbook useful primarily as a guide, and may be able to eliminate some of the steps outlined.

PDS has developed requirements and standards for data sets to help ensure that the data it provides to users in the science community are complete, accurate, and easily accessible. PDS works with flight projects during all phases of the mission; before launch to help flight project personnel design data sets and data production systems, during the active mission to archive new data sets, and after the active mission to archive past data sets or newly derived data sets. PDS also works with non-flight project personnel to restore old data sets and observations of interest to planetary scientists. This manual is intended for all types of data suppliers and developers working with PDS.

## 1.3 Document Scope

The information included here describes Version 3.1 of the Planetary Data System data preparation procedures for preparing data sets that meet PDS archive standards. These procedures, in conjunction with the standards provided in the PDS Standards Reference, supercede information given in the previous versions of this document.

## 1.4 Document Overview

This workbook is organized by the steps involved in the data preparation and submission process. The requirements, standards, and procedures presented here reflect the most recent PDS updates. Checklists are provided at the start of Chapters 5-10 outlining the steps described in the chapter.

Chapter 2 - provides the rationale for submitting data to PDS
Chapter 3 - describes what is meant by a PDS data set
Chapter 4 - provides an overview of the data preparation and submission process
Chapter 5 - provides information on establishing contact with PDS
Chapter 6 - discusses steps in archive planning
Chapter 7 - discusses steps in archive design
Chapter 8 - discusses steps in assembling and validating data sets
Chapter 9 - discusses PDS review procedures
Chapter 10 - discusses steps for archiving and distributing PDS data sets

Appendix A - provides PDS contact information
Appendix B - provides information on PDS Tools
Appendix C - provides a glossary of PDS terms
Appendix D - provides a list of acronyms

## 1.5 Related Documents

The latest version of PDS standards documents may be accessed electronically from the PDS World Wide Web homepage-- URL http://stardust.jpl.nasa.gov/ . Hardcopies of these documents may also be obtained from the PDS Operator (see Appendix A - Whom to Contact).

**Planetary Science Data Dictionary**, JPL D-7116, Rev. C, November 20, 1992. - A dictionary of standard terms used to describe planetary data products. This document is required for preparing data sets for archiving with PDS.

**PDS Standards Reference**, JPL D-7669, part 2, Version 3.1, August 3, 1994. - Detailed standards and examples for defining data sets for archive, organizing these on physical volumes, and labelling individual data products. This document is required for preparing data sets for archiving with PDS.

The following documents are available as hardcopy only from the PDS Operator.
**PDS Toolbox Overview**, Toolbox Release 4, JPL D-10263, November 16, 1992. - An overview of PDS supported software libraries and utilities and available User's Guides. These are optional documents that may help the data producer in a variety of steps during the overall data preparation process. Additional information about PDS Tools is provided in Appendix B.
**PDS Label Verifier User's Guide**, Version 1.2, JPL D-8923, November 9, 1992.
**User's Guide for the PDS Label Generators**, JPL D-10265, November 2, 1992.
**User's Guide for the PDS Simple Label Editor**, Release 2.4, JPL D-9436, November 1, 1992.
**PDS Table Browser User's Guide**, JPL D-10264, October 15, 1992.
**PDS Toolbox Utilities User's Guide**,Version 1.1, JPL D-9420, October 28, 1992.
**PDS Label Library User's Guide**, Version 1.1, JPL D-8922, March 31,1992.

## 1.6 Overview of PDS

PDS was created by NASA to provide a cost-effective archive system for planetary data from past, present and future planetary missions. Sponsored by NASA's Solar System Exploration Division (Code SL), PDS provides a system of organization and support that is designed to stimulate research, facilitate data access, and support correlative analysis. The PDS is based upon a widely distributed, electronically connected architecture which allows data sets to remain in the user community where they are most likely to be used and perpetuated, and where scientific expertise is available to assist the research scientist. PDS is divided into several operational units (also known as Nodes) shown schematically in Figure 1.1. A Node is a service and research center funded for operation by the PDS.

### 1.6.1 The Central Node

The Central Node provides overall project management for PDS, providing coordination and support services for the Discipline Nodes. The Central Node is also responsible for the maintenance and distribution of data archive standards, and evaluates emerging archive technologies. It provides an interface to active planetary missions, provides software for preparing and validating data, and maintains a central catalog system. The Central Node resides at the Jet Propulsion Laboratory (JPL) in Pasadena, California.

The PDS Central Data Set Catalog is available to help a user locate and order complete data sets or volumes of interest. This general catalog provides high level information about various missions, spacecraft and earth based instrument hosts, instruments, planetary targets, and associated sets of data (data sets) that are available from PDS. Once a data set of interest has been identified by a user, connections to more detailed information about portions of the data set (data products) may also be available. These more detailed descriptions of data products are kept by PDS in Data Product Catalogs, located and maintained by various PDS Discipline Nodes. For further information, contact the PDS Operator listed in Appendix A.

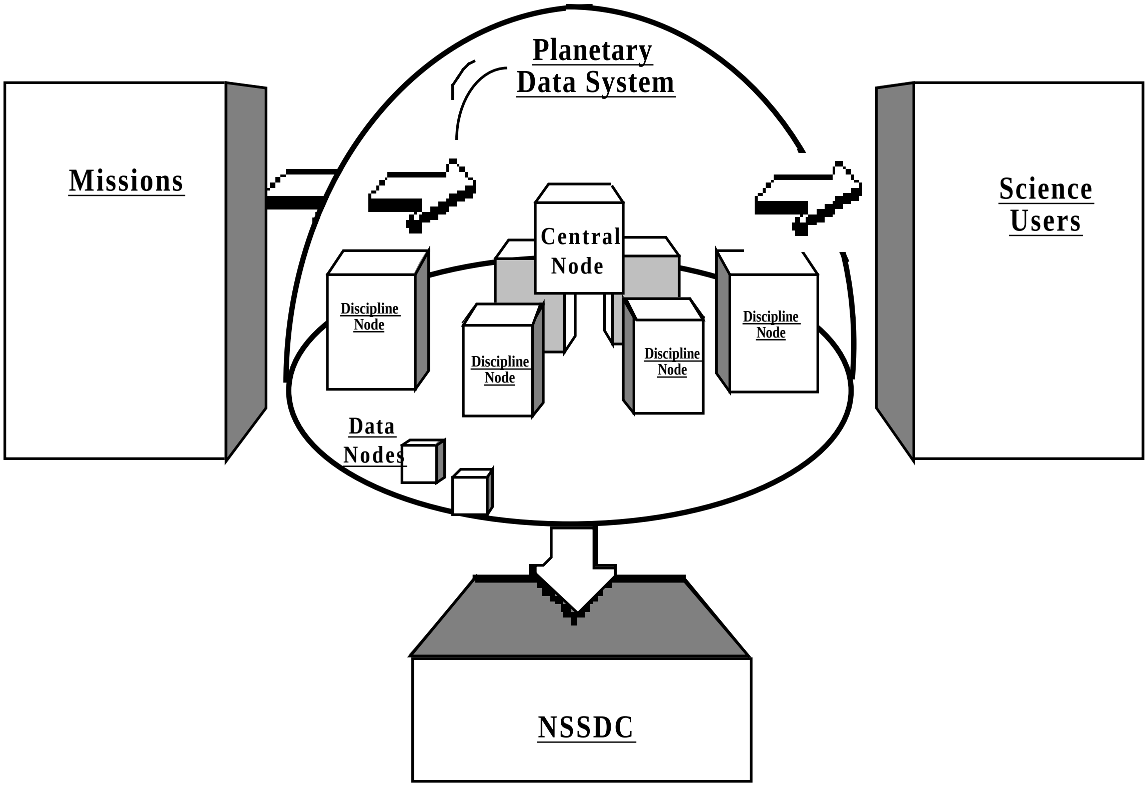



Figure 1.1 Schematic View of PDS

### 1.6.2 The Discipline Nodes

PDS includes seven Discipline Nodes located at institutions around the country, each having expertise in a particular area of planetary science. These Nodes are listed below (see Appendix A, Whom to Contact, for additional information):

Planetary Atmosphere (University of Colorado)
Small Bodies (University of Maryland)
Geosciences (Washington University, St. Louis)
Imaging (U.S. Geological Survey, Flagstaff, and JPL)
Rings (Ames Research Center)
Planetary Plasma Interactions (UCLA)
Navigation and Ancillary Information Facility (JPL)

Each of the Discipline Nodes is responsible for archiving data sets specific to its area of expertise and distributing these data to users in the science community. PDS Discipline Nodes work with other scientists to prepare data collected or derived from a wide variety of planetary missions and observations, and are involved in the planning and design of new data sets for active flight projects.

There are several subnodes formally associated with PDS Discipline Nodes, each contributing a special area of scientific expertise to PDS. Discipline Nodes may also establish Data Nodes for the purpose of restoring a data set from a past planetary mission or making significant improvements to a recently obtained data set. These Data Nodes are dynamic in nature, and exist only for the duration of a specific task.

PDS Discipline Nodes also develop and make available methodologies and software for examining specific types of data. Several Discipline Nodes provide electronic access to on-line data holdings and provide ordering mechanisms for receiving subsets of PDS data sets. In additon, detailed product catalogs are maintained by several of the Discipline Nodes, together with software for selecting specific products of interest, displaying selected products, and performing various types of analyses.

**Planetary Atmospheres Node**

The Planetary Atmospheres Node is a consortium of research groups led by the University of Colorado’s Laboratory for Atmospheric and Space Physics (LASP). Researchers associated with the Node curate data sets within their subdisciplines and provide a wealth of research expertise to space scientists and other users of the PDS. The various groups making up the Node, their areas of scientific expertise, and the manager of each Subnode, are:

University of Colorado
UV spectroscopy, aeronomy, radiative transfer, microwave spectroscopy, surface-atmosphere interactions, cometary
atmospheres, data management
Ames Research Center
Atmospheric dynamics, atmospheric modeling
University of Arizona
UV and EUV spectroscopy
Goddard Space Flight Center
IR radiometry, IR spectroscopy
New Mexico State University
Imaging of planetary atmospheres, climatic data from long-term monitoring

**Small Bodies Node**

The Small Bodies Node archives both spacecraft data and ground-based observations of comets, asteriods and dust. Managed by the University of Maryland, College Park, the node consists of three subnodes in the United States and a European subnode:

University of Arizona - Interplanetary Dust
University of Hawaii - Asteroids
University of Maryland, College Park - Comets
Konkoly Observatory, Budapest Hungary

**Geosciences Node**

The Geosciences Node is responsible for working with planetary missions to help ensure that data of relevance to geoscience disciplines are properly documented and archived with PDS. Managed by the Earth and Planetary Remote Sensing Laboratory at Washington University, geoscience subnodes and their specialties include:

Washington University - Geophysics
Massachusetts Institute of Technology - Microwaves
Arizona State University - Spaceborne Thermal Data
Brown University - Spectroscopy

**Imaging Node**

The Imaging Node is responsible for archiving NASA's primary digital image collections from past, present, and future planetary missions. The node provides to the NASA planetary science community the digital image archives, necessary ancillary data sets, software tools, and technical expertise necessary to fully utilize the vast collection of digital planetary imaging. Managed by the United States Geological Survey in Flagstaff, Arizona, the Imaging Node includes a subnode at the Jet Propulsion Laboratory.

**Planetary Rings Node**

The Planetary Rings Node is responsible for restoring, archiving, and publishing data sets describing planetary ring systems. This includes all the relevent data sets acquired from previous and future spacecraft, as well as Earth-based observations. The Rings Node is a cooperative project of Ames Research Center and the Center for Radar Astronomy at Stanford University.

**Planetary Plasma Interactions Node**

The Planetary Plasma Interactions (PPI) Node is responsible for the acquisition, preservation, and distribution of fields and particle data from all the planetary missions (excluding most Earth observations). The PPI node is managed by the University of California, Los Angeles. Subnodes and their specialties include:

University of California, Los Angeles - Inner Planets
University of Iowa - Outer Planets
Planetary Radio Science

Navigation and Ancillary Information Facility

The Navigation and Ancillary Information Facility (NAIF) is responsible for development and implementation of the SPICE concept - a means for providing ancillary observation geometry and related data useful for planning observations of solar system targets, and for interpretation of science instrument data sets returned from planetary spacecraft.

SPICE data sets, called kernels, exist for spacecraft trajectory (S); planet and satellite ephemerides and associated physical and cartographic constants (P); instrument descriptions (I); orientation as a function of time of spacecraft structures (C); and spacecraft and ground system events (E). NAIF archives and distributes SPICE kernels produced by a number of flight projects accompanied by an extensive portable software toolkit for use with the kernels.

Data Distribution Laboratory

The Data Distribution Laboratory (DDL) was created as a part of an effort to aid scientists in the archive and distribution of science data sets using Compact Disk-Read Only Memory (CD-ROM) technology. The DDL provides scientists with the ability to premaster and cut prototype CD-ROMs on-site. The entire CD-ROM production process is supported, from conversion of existing data to CD-ROM formats, to the actual production of a premastered tape ready for shipment to an outside disk mastering facility. The DDL is located at the Jet Propulsion Laboratory.

### 1.6.3 The National Space Science Data Center (NSSDC)

The National Space Science Data Center (NSSDC) is the long term archive for all space science data. PDS serves as the primary entry point for digital planetary data to the NSSDC. PDS prepares new digital data sets from on-going planetary missions, and restores data sets from previous missions, validating the data quality and entering information about the data sets and data products into PDS catalogs. In addition, information on the data sets is provided to the NSSDC which is used in the NSSDC catalogs, including the Master Directory (MD). The NSSDC is the sole archive for non-digital data sets such as film and paper products. The NSSDC also distributes bulk orders for digital data sets and all orders for non-digital data.

# Chapter 2

## Why Submit Data to PDS?

Establishing a data archive such as the NSSDC is only the first step in making data useful over an extended period of time. Scientists not immediately familiar with the archive system must be able to locate data sets of potential interest and be able to extract meaningful subsets of data for further study. With this in mind, PDS has established:

- Standards for identifying and organizing planetary data sets.
- Common nomenclature and data structures especially suitable for archiving planetary data.
- A catalog system which allows locating and searching of data, and an ordering system which facilitates rapid distribution of data to the planetary community.
- Software tools for manipulating data stored in standard structures.
- Data management expertise which is available to both flight projects and individual researchers interested in archiving planetary data.

PDS seeks not only to preserve planetary data that might otherwise be lost or poorly documented, but to make it readily accessible to an ever-increasing community of potential users supported by state-of-the-art technology.

If you are interested in learning more about how to use PDS resources or obtain PDS data, contact the PDS Operator listed in Appendix A. If you want to know more about the PDS concept of a data set, it’s time to move on to Chapter 3.

# Chapter 3

# What is a Data Set?

One of the objectives of the PDS is to introduce consistency in the contents and organization of planetary data sets. The PDS has introduced the concept that an *archive quality data set* must include everything that is needed to understand and utilize the data. Towards this goal, the PDS has worked with the NSSDC, PDS Discipline Nodes, numerous flight projects, and individual scientists and programmers to develop approached to ensure that this consistency is achieved.

The remainder of this workbook focuses on the process of preparing and submitting data to PDS. To do this effectively, you must first understand the terminology used within the text to describe different levels and groupings of data, as well as the terminology used to describe its physical organization.

Section 3.1 describes the components of a data set. Section 3.2 then describes how these data set components can be organized onto digital media.

## 3.1 What Constitutes a Data Set?

The organization of planetary science data can be quite complex. Producing an archive quality data set means including everything that is needed to understand and utilize data. The relationship between the data and everything else can be complex also. In addition, data sets may be grouped together into larger entities to serve a particular scientific objective.

To distinguish between different levels of data, groupings of data, and relationships between data and everything else, the following terminology will be used. Figure 3.1 shows the relationship between these entities.

**Data set collection** - A data set collection consists of data sets which are related by observation type, discipline, target, or time, for a specific scientific purpose. The collection of data sets is treated as a unit and are catalogued and ordered as a unit.

**Data set** - The accumulation of data products, supplemental data, software, and documentation, that will completely document and support the use of those data products. A data set can be part of a data set collection. A data set may include the following components:

**Documentation** - Textual material which describes the mission, spacecraft, instrument, and data set. This can include references to science papers, calibration reports, and other text useful for interpreting the data.

**Catalog Data** - Descriptive information about a data set (e.g., mission description, instrument host description, instrument description), expressed in the Object Description Language (ODL) which is suitable for loading into a catalog.

**Software** - Software libraries, utilities, or application programs to access/process the data products.

**Calibration data** - Calibration files used in the processing of the raw data or needed to use the data.

**Geometry data** - Relevant files (e.g., SEDRs, SPICE kernels) needed to describe the observation geometry.

**Indices** - Information which allows the user to locate the data products of interest, such as a table relating latitude/longitude ranges to file names.

**Data Products** - Labeled groupings of data resulting from a scientific observation. A product label expressed in ODL, identifies, describes, and defines the structure of the data. Examples of a data product are planetary images, spectrum tables, and time series tables. A data product is a component of a data set. Each data product is made up of the following:

**PDS Product Labels** - A label expressed in ODL which identifies, describes, and defines the structure of the data.

**Primary data objects** - A grouping of data resulting from a scientific observation. The actual science data, such as an image or table, representing the measured instrument parameters. A data set usually contains many data objects. For example, there are 10,000 images in the Voyager Neptune imaging data set.

**Secondary data objects** - Any data needed for processing or correctly interpreting the primary data object. Each data object may have secondary data objects associated with it, such as a histogram (derived from the primary data object) for each image in a series of related images.

## 3.2 How Data Sets are Physically Organized

The data set components identified in Section 3.1 must be organized onto digital media for archiving and distribution. Small data sets that are related can be grouped onto a single volume. For example, a set of 22 small data sets from six fields and particles instruments are archived on a single volume for the Voyager 2 Neptune encounter. Frequently, because of the size of data sets, multiple volumes are required.

Figure 3.2 shows how many data sets are physically organized. A volume represents a single unit of media. The archive media supported by PDS are CD-ROMs and magnetic tape, as discussed in the *Media Formats for Data Submission and Archive* chapter of the PDS Standards Reference. A volume set represents a collections of volumes organized by a common attribute.

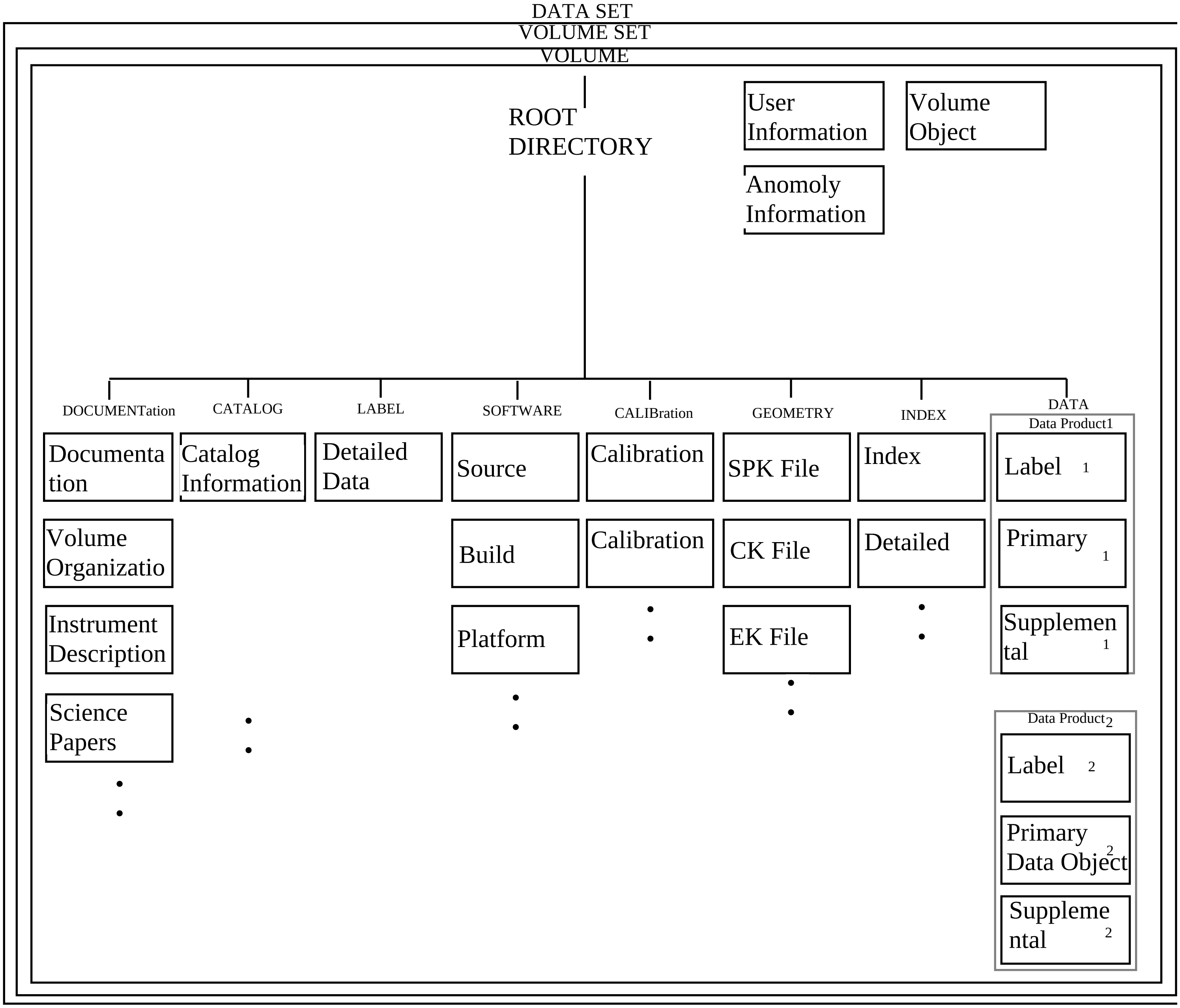



Figure 3.1 Components of a Data Set

Within each volume is a directory structure, listing the subdirectories and files contained on that volume. Magnetic tapes have a “virtual” directory structure provided in directory and file objects included on the volume. The files shown in Figure 3.2, between directory and subdirectory, represent the existence of special files (AAREADME.TXT, VOLDESC.CAT and optional ERRATA.TXT) that are located only at the root directory.

There are seven standard subdirectories of each volume to accommodate the documentation, software, ancillary data, and support data for data sets. The names of these subdirectories are:

1) DOCUMENTation
2) CATALOG
3) LABELs
4) SOFTWARE
5) CALIBration data
6) GEOMETRY data
7) INDEX

The eighth and additional subdirectories are the DATA subdirectories containing files of data products. The DATA subdirectories are named as described in the *Directory Types and Naming* chapter of the PDS Standards Reference. DATA subdirectories may be nested to the limit of the media. Each file in the DATA subdirectories contains a data product consisting of a PDS label and the actual data.

The physical organization shown in Figure 3.2 can vary slightly depending upon the structure of your data set. Alternate volume organizations are discussed in detail in the *Volume Organization and Naming* chapter of the PDS Standards Reference. Examples are:

- Multiple data sets can be contained on a single volume.

- More complex data sets or volumes with multiple level of dependencies can be organized in a logical hierarchy.

- A data set can reside on several volumes. Figure 3.3 illustrates the physical organization of the Voyager Jupiter Images Data Set residing on 3 CD-ROMs.

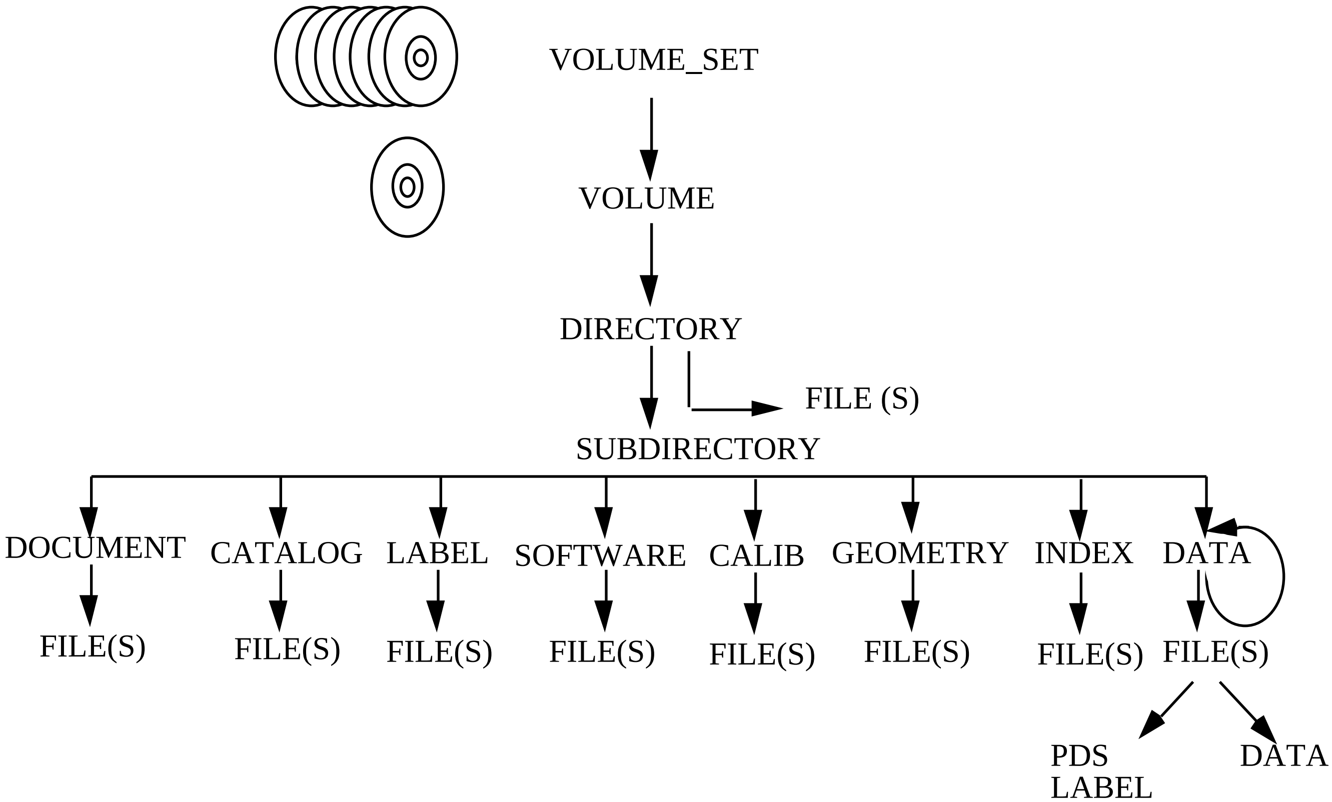



Figure 3.2 Physical Organization of a Data Set

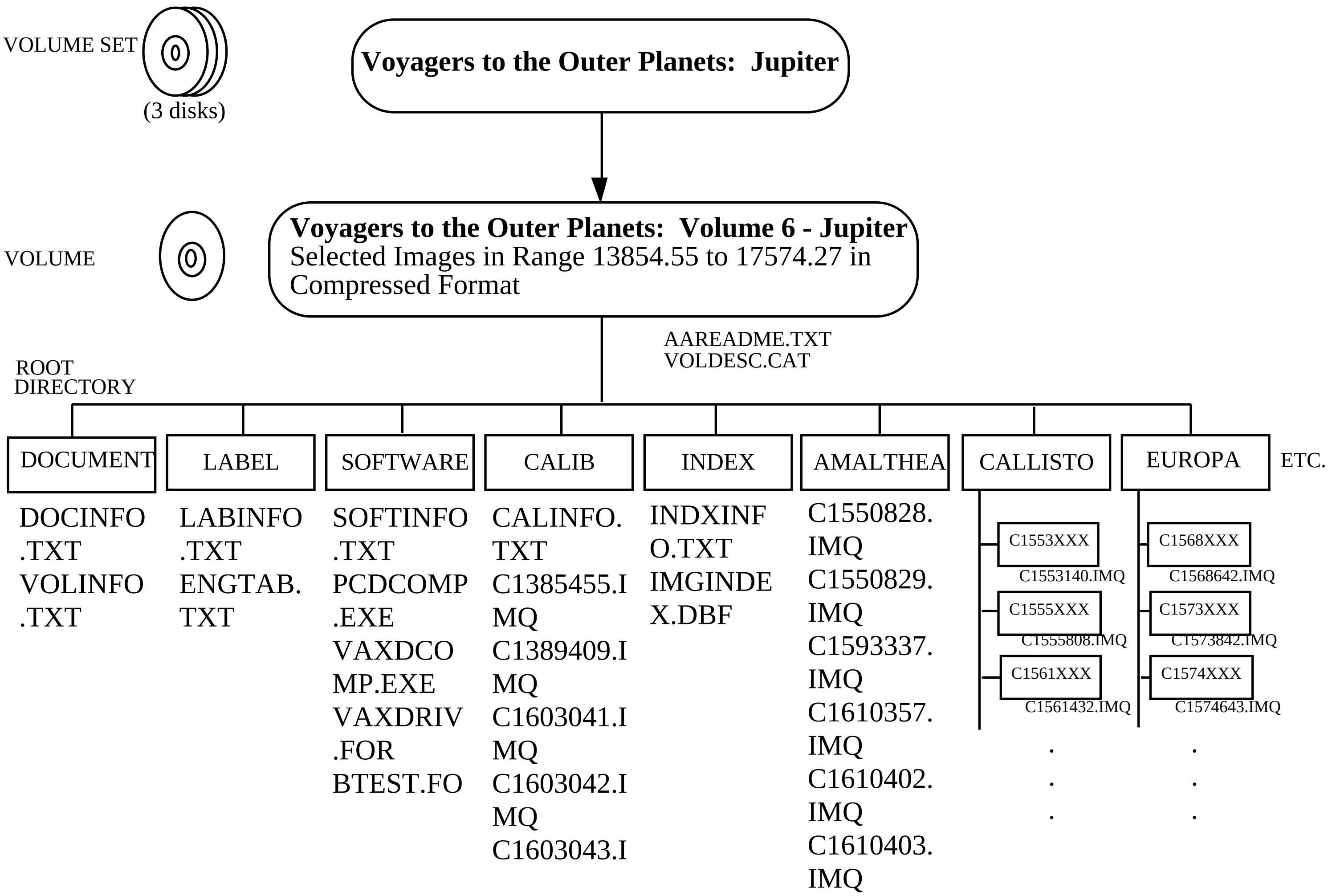



Figure 3.3 Example of a Data Set

Using Figure 3.3 as an example, we can review the terminology used within this workbook as it applies to an actual data set, and take note of how the preparers of this data set decided to organize their data products.

- The Jupiter Selected Image data set is a part of the series “Voyagers to the Outer Planets”, a collection of Voyager 1 and 2 images of the outer planets.
- The Jupiter Selected Image data set resides on a volume set of three disks, Volume 6 through 8. Only the organization for Volume 6 is shown in Figure 3.3. The directory structure for Volumes 7 and 8 is similar.
- The files within each subdirectory are listed below the directory name. Note that the files AAREADME.TXT, and VOLDESC.CAT reside at the root directory level. Also note that there are subdirectories beneath the CALLISTO and EUROPA data directories.
- The data preparers for the Voyager Jupiter Selected Image Data Set chose to organize the data directories by target (i.e., Jupiter’s moons Amalthea, Callisto, and Europa are shown above, Ganymede, Io, and Jupiter are also on the disk). The data preparers also decided to make an image frame (800 x 800 pixels) the primary data, each image frame corresponding to one file. The primary data objects are identified by spacecraft clock time (FDS count) and these clock times are used to name the data files (with the .IMQ file extension). In addition to the primary data object, two secondary data objects are included in each file. These are an image histogram and an engineering table.

Table 3.1 associates PDS data set terminology with this data set example.

| Term | Example |
|---|---|
| Data Set | Voyager Jupiter Selected Images Data Set |
| Volume Set | Volumes 6 through 8 |
| Volume | Volume 6 |
| Data Product | Image of Jupiter’s moon, Europa |
| Primary Data Object | Image |
| Secondary Data Objects | Histogram, engineering table |

**Table 3.1: Using the Terminology**

You should now have an idea of what a data set is and some of the decisions you must make during the data preparation and submission process.

# Chapter 4

# The Data Preparation Process

Once you’ve decided to archive your data set with PDS, you’ll discover there are six steps in the archiving process:

1) orientation - finding out what PDS will expect
2) archive planning - deciding what to archive, when, and generally how
3) archive design - learning the details of putting an archive data set together
4) data set assembly and validation - pulling the pieces together
5) data set reviews - the final PDS quality check
6) delivery - passing the result to PDS

Subsequent chapters of this workbook provide detailed information for completing each of these steps.These steps to submit planetary science data to the PDS are similar whether data are being submitted from an active flight project is being restored from past projects or observations.

The approach to archiving will vary depending on your relationship to PDS. For active flight projects, early and frequent contact is desirable and will often be formalized. Figure 4.1 shows a typical flight project time-line and the correlation between project and PDS events. PDS personnel at the Central Node and one or more Discipline Nodes may become involved, providing a wide range of support. This will help to ensure that the enormous volume of data expected later flows smoothly into the system. Cooperative development along these lines is cost-effective for both the flight project and PDS.

For data restoration or for archiving simple (possibly on-going) observations from a single source, liason with the PDS Discipline Node most appropriate for your specialty may be sufficient. In many cases, your Discipline Node associates may perform many of the steps identified.

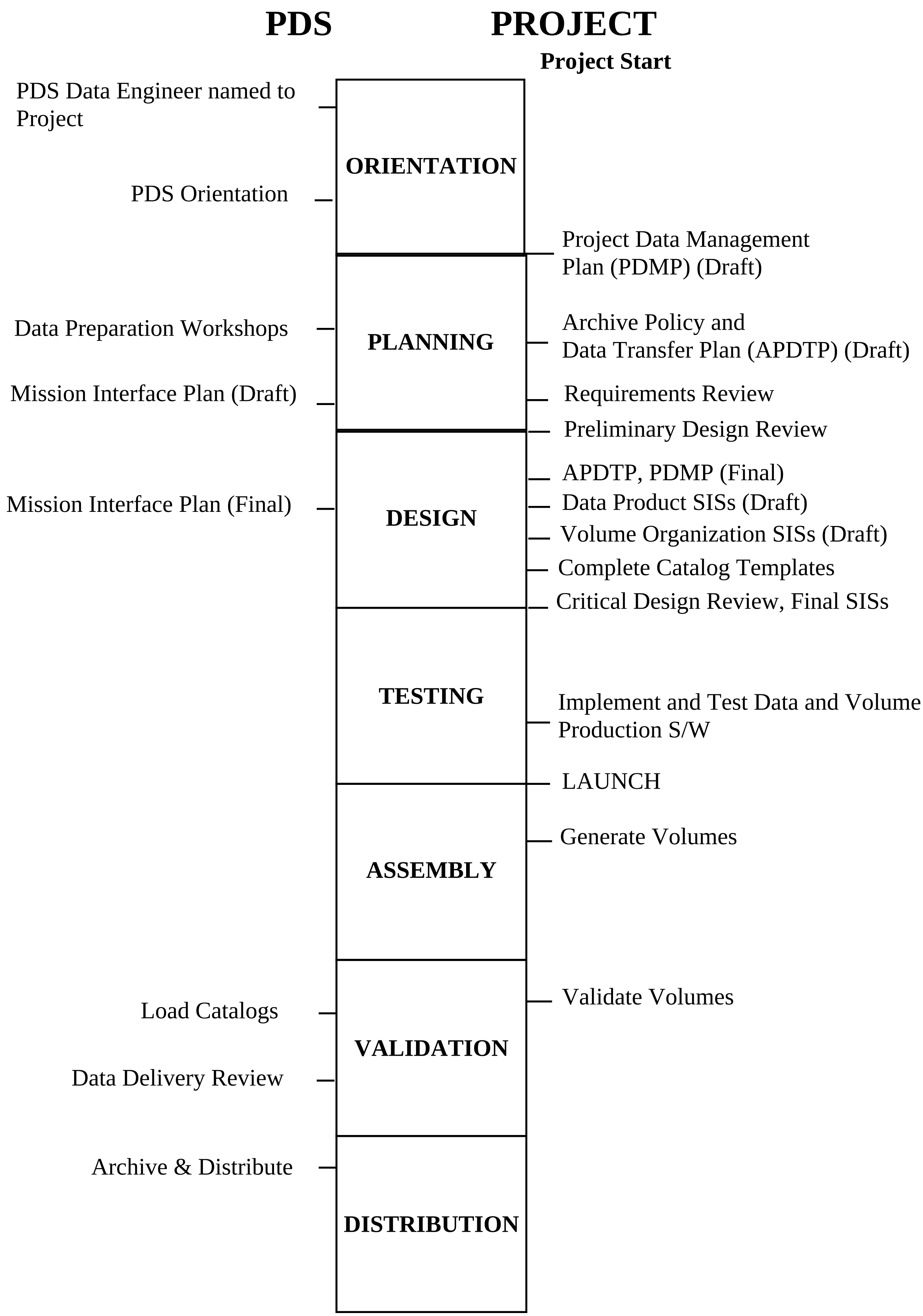



Figure 4.1 PDS / Project Time-line

# Chapter 5

## Orientation Checklist

# Chapter 5

# Orientation

This section describes the first phase of the archive process. During this phase contact is established between PDS and flight project managers or potential data suppliers.

## 5.1 Establish Contact with PDS

The first point of contact with the PDS should be through the PDS Mission Products Manager for active flight projects or through an appropriate Discipline Node Manager. (see Appendix A, Whom to Contact). The Mission Products Manager coordinates support for all aspects of archiving active flight project data and provides additional support to PDS Discipline Nodes for help with data restoration activities.

## 5.2 Provide General Information to PDS

For active flight projects, the following general kinds of information will be helpful during early contact with the PDS Mission Products Manager:

- General project information ( mission description, types and numbers of instruments, science objectives, estimated volume of data to be archived)
- Major project milestones (launch date, encounter date, major project reviews)
- Project contact(s) for archive issues

For data restorations or other kinds of experiment or observation data, general information that may be useful to the PDS Discipline Node Manager, to the extent it is known, may include:

- Scientific value of the data
- Descriptive information about the data (spacecraft, instrument/experiment, targets, events, range of data)
- Physical description of the data (current media, total data volume, data organization, formats, data condition)
- Description of supporting software needed for understanding and/or analyzing the data (e.g. data processing, reduction, analysis software, supporting documentation)
- Description of relevant ancillary data (SEDRs, SPICE kernels, calibration data, etc.)
- Special considerations (e.g., tapes in danger of deterioration, PIs may move on to other projects, data is in demand, data would be easy to restore)

## 5.3 Obtain PDS Orientation Material

For active flight projects, a PDS Orientation will be scheduled. At this orientation, you will be given an overview of PDS, including general descriptions of the roles and responsibilities of the project, the PDS Central and Discipline Nodes and the NSSDC during the archive process. A Data Preparation Package will be handed out, providing you with additional information.

You may also arrange for an overview and demonstration of PDS software tools to help determine their applicablilty to your project needs. These demonstrations are geared to software or system engineers involved in designing a ground data system for the project. Further information on PDS Tools is provided in Appendix B.

For data restorations, if you will be preparing the data for archive yourself, your Discipline Node contact may suggest that you obtain a Data Preparation Package also.

The Data Preparation Package consists of:

PDS Data Preparation Workbook (this document)
PDS Standards Reference
Planetary Science Data Dictionary
PDS Toolbox Overview

## 5.4 Establish Technical Contacts

For flight projects, the Mission Products Manager will identify a data engineer from the Central Node and one or more PDS Discipline Nodes who will provide technical support to the project. The project should also identify their management and technical contacts for archive issues at this time.

Generally, flight projects will establish teams (such as the Data Product Working Group on Magellan or the Data Working Group on Galileo) to address project-wide archive issues. These teams are chaired by members of the project's science community and consist of representatives from each instrument team, PDS representatives, and flight project engineers involved in data archiving.

In addition, PDS may establish a team, called a Mission Interface Team (MIFT), which meets more frequently than the flight project team, and addresses some of the more detailed archive issues. The MIFT will consist of members of the PDS Central and Discipline Nodes, the NSSDC, and the flight project.

For data restorations, your primary technical contact will be with the PDS Discipline Node. In addition, the PDS Mission Products Manager will identify a Data Engineer from the Central Node to assist the Discipline Node as needed during the archive process.

# Chapter 6

## Archive Planning Checklist

- ☐ 6.1 Prepare a Project Data Management Plan (PDMP)
- ☐ 6.2 Prepare an Archive Policy and Data Transfer Plan (APDTP)
- ☐ 6.3 Plan for Updates to the PDMP and APDTP
- ☐ 6.4 Keep the PDS Central Node Data Engineer Informed
- ☐ 6.5 Participate in Planning Meetings
- ☐ 6.6 Review/Sign off Mission Interface Plan

# Chapter 6

# Archive Planning

For active flight projects, archive planning consists of identifying the data to be archived, developing a detailed archiving schedule, and defining an end-to-end data flow through the ground system. Part of this planning also defines roles and responsibilities of the variety of teams involved in producing final archive products.

This activity is less formal for data restorations, and does not require preparation of the documents discussed in the following steps.

## 6.1 Prepare a Project Data Management Plan (PDMP)

A Project Data Management Plan (PDMP) is required by NASA for all new projects. This plan provides a general description of the project data processing, cataloging, and communication plan. The PDS Project Manager is a signatory on this document.

PDS will provide assistance in developing the PDMP at the request of the project. A copy of the PDS Guidelines for Project Data Management Plans, JPL Document D-5111 and a sample PDMP may be obtained from the PDS Mission Products Manager. Once the project has completed a draft of this document, PDS and the NSSDC will participate in it’s review, and help to identify and resolve archive related issues.

## 6.2 Prepare an Archive Policy and Data Transfer Plan (APDTP)

The Archive Policy and Data Transfer Plan (APDTP) provides a detailed description of the production and delivery plans for archive products for a project. The PDS Project Manager is a signatory on this document.

The contents of an APDTP include:

1) a list of proposed data sets for archive, including a general description of the science content of each data set and a prioritization of data sets in terms of expected general science popularity for each data set.

1) a schedule of availability of data sets, keeping in mind:
- in general, PDS cannot accept proprietary data.
-project users of PDS data do not have priority over other PDS users even during an active mission. The project may want to keep its own copies of certain data sets for a period of time.
-all products do not need to have the same transfer schedule.

1) the storage medium, count, total size, processing level, etc., of each proposed data set.

1) proposed data transfer mechanism.

PDS will also provide assistance in developing the APDTP at the request of the project. Example APDTPs may be obtained from the PDS Mission Products Manager or from the Central Node Data Engineer assigned to the project. PDS and the NSSDC will participate in reviewing the APDTP,

## 6.3 Plan for Updates to the PDMP and APDTP

It is inevitable that changes will occur that will affect both the PDMP and the APDTP. In particular, detailed data set lists and schedules found in the APDTP appendices often develop over time, making these appendices “working” guidelines for archive planning. The PDS Central Node Data Engineer should be notified of changes to these plans as they occur, and document revisions should be scheduled periodically. Changes include additions and deletions of products, changes in schedule, and changes in quantity, product content or format.

## 6.4 Keep the PDS Central Node Data Engineer Informed

For active flight projects, the assigned PDS Central Node Data Engineer will need to be placed on all relevant project distribution lists. You may also want to invite this PDS representative to attend certain project meetings that involve discussions about archiving. By including PDS in ground data system planning efforts, there may be ways to reduce the work required later for preparing data for archiving.

## 6.5 Participate in Planning Meetings

For active flight projects, the PDS Central Node Data Engineer or Mission Interface Team (MIFT) leader will schedule regular planning meetings. During the pre-launch phases of a project, these meetings may be used to help with drafting and reviewing the PDMP and APDTP project documents and defining project interfaces with PDS. After launch, when the emphasis shifts to archive testing and production, these meetings may be used to develop details of archive transfer procedures between the project, PDS, and the NSSDC.

For data restorations, informal planning meetings may be held to coordinate the transfer of data as well as steps to discuss details of the actual restoration work. These will typically be organized by the Discipline Node involved with the restoration.

## 6.6 Review/Sign off Mission Interface Plan

For active flight projects, a Mission Interface Plan will be written by the PDS Central Node Data Engineer or another member of the Mission Interface Team. The Mission Interface Plan establishes the general roles and responsibilities of PDS Central and Discipline Nodes, the NSSDC, and each project team that has an interface with PDS. There will be signatories on this document for each identified interface. This document will require review and signature, and issues may be brought to MIFT planning meetings for discussion.

# Chapter 7

# Archive Design Checklist

- ☐ **7.1 Review PDS Standards**

- ☐ **7.2 Design Data Products and Labels**
- ☐ 7.2.1 Define the Data Product
- ☐ 7.2.2 Estimate File Sizes
- ☐ 7.2.3 Determine the Data Format
- ☐ 7.2.4 Determine the Data Objects
- ☐ 7.2.5 Determine Data Product File Configurations
- ☐ 7.2.6 Design Data Product Labels
- ☐ 7.2.7 Validate Label Syntax
- ☐ 7.2.8 Validate Data Object Definitions
- ☐ 7.2.9 Write Data Products SISs

- ☐ **7.3 Design the Data Set or Collection**
- ☐ 7.3.1 Define the Purpose and Scope of the Data Set or Collection
- ☐ 7.3.2 Determine Other Data Set or Collection Components
- ☐ 7.3.3 Create Data Set or Collection Names and Identifiers

- ☐ **7.4 Determine the Storage Medium**

- ☐ **7.5 Design Volume Set/Volume**
- ☐ 7.5.1 Map Data Sets or Collections to Volumes
- ☐ 7.5.2 Name the Volume Set
- ☐ 7.5.3 Name the Volumes
- ☐ 7.5.4 Determine non-data Subdirectories and Files for each Volume
- ☐ 7.5.5 Determine the Data Organization
- ☐ 7.5.6 Establish Data Product File Naming Convention
- ☐ 7.5.7 Establish Directory Naming Convention
- ☐ 7.5.8 Determine the Indices Needed
- ☐ 7.5.9 Write Volume Organization Software Interface Specifications (SIS)

- ☐ **7.6 Design Data Production Process**
- ☐ 7.6.1 Design Label Generation Process
- ☐ 7.6.2 Design the Volume Production Process

- ☐ **7.7 Plan Data Validation Process**
- ☐ 7.7.1 Investigate Validation Tools
- ☐ 7.7.2 Write Data Validation Plan
- ☐ 7.7.3 Design, Implement, and Test Project Data Validation Software
- ☐ 7.7.4 Plan for updates to Data Validation Procedures and Software

- ☐ **7.8 Complete the Catalog Object Templates**
- ☐ 7.8.1 Obtain the Catalog Object Templates
- ☐ 7.8.2 Review the structure of the Catalog Object Templates
- ☐ 7.8.3 Complete the Catalog Object Templates
- ☐ 7.8.4 Validate the Completed Catalog Objects
- ☐ 7.8.5 Submit the Completed Catalog Objects
- ☐ 7.8.6 Plan for updates to the Templates

# Chapter 7

# Archive Design

Archive design consists of:

- a review of PDS standards
- data product and label design
- data set or collection design
- determination of storage medium
- volume set/volume design
- data production process design
- data validation planning
- catalog object template completion

This chapter describes each of these activities. These tasks are not meant to be sequential. In many cases there may be several iterations between various steps. If you are involved in designing an archive for a flight project or as a data node for a data restoration, you may be performing many of these activities yourself with the help of your PDS contacts. For PDS initiated data restorations, most of this design work is done by the responsible Discipline Node.

## 7.1 Review PDS Standards

Once a data set has been identified for archiving, it is time to review various ways to organize the data so that it will be most accessable and usable to a broad community. PDS has developed standards to help you in both organizing and describing your data. A brief review of this Data Preparation Workbook, the Standards Reference, and the Planetary Science Data Dictionary (PSDD) should provide you with a basic understanding of what PDS standards are all about. It helps to understand that both the Standards Reference and the PSDD are truely references, and will answer many questions that come up as part of the archive design process.

For active flight projects, PDS usually sponsors one of more Data Preparation Workshops that will focus on the use of these archive standards. For both data restorations and flight projects, the PDS Data Engineer and Discipline Node contacts you have made are available to answer your questions and provide guidance to you on archive design issues. They will be able to provide you with examples of data volumes already archived with PDS, so you can see just how the data can be packaged, labeled, documented, catalogued, and ordered. Usually, seeing examples of what has been done in the past is the easiest way to understand what you will be designing.

Don’t hesitate to ask for help! The standards documents are updated regularly, and it is the goal of PDS to try to make these easy to follow. If you are having trouble understanding them, then PDS wants to try to correct this. Your suggestions are appreciated.

## 7.2 Design Data Products and Labels

This activity includes the determination of both the contents and the file format of the primary data product for a data set. This includes the definition of the data objects that make up the product and the definition of the data product label. For active flight projects, a Data Product Software Interface Specification (SIS) is often required to document this design.

### 7.2.1 Define the Data Product

This activity involves the definition of the data product from a scientific perspective. This includes defining how the data should be divided into individual data products and determining which parameters or measurements will be included. Sometimes the definition of a data product may seem obvious. For example, most data sets composed of images from a particular spacecraft have defined a data product as a single image. Tabular data sampled and organized by observation time may be packaged into data products containing an hour’s worth of data, for example.
Typically, the definition of a data product will depend on several factors. The most important consideration, however, is the probable way the data will need to be accessed and the expected frequency of access. There will probably be some iteration in defining the data product, and the next step, estimating the file sizes.

### 7.2.2 Estimate File Sizes

Once you’ve identified a likely data product, make some estimates of the size of data files that will result from this definition. Be sure to take into consideration variable data rates that may apply to the data set and determine both average size files and maximum size files that could be received.

Data products should be designed so that the individual file sizes are reasonable for the expected use. It would be a mistake, for example, to design a data product with file sizes of 60 megabytes if you expect the data to be used with a personal computer. However, this may be entirely adequate for low-level products that will be processed infrequently and are geared toward a workstation environment.

There is also a trade-off between defining data products that result in large file sizes and definitions that result in large numbers of files to manage in a data set. If the data products are expected to be made available on-line and transferred electronically, very large files (over 10 megabytes) may not transfer as reliably as files that are 1 megabyte or less. If typical requests will require transfers of many small files, then the record keeping, transfer, and handling may be more complex.

### 7.2.3 Determine the Data Format

The format for each type of data product file that will be present in the data set or collection is determined next.

When restoring data sets, this may require some research, since documentation of data formats is sometimes discarded after the initial set of analysis software is written or is not updated when formats are altered. Data formats for existing data may be found in software user's guides, requirements or interface specification documents, or embedded in the software source code designed to read or write the data. In many cases, it may be best to reformat the data into a more portable format,

Once again, the anticipated use of the data will have a great bearing on the type of format and organization of the data. Generally, the more structured the anticipated usage, the less concern about the storage format. For example, compressed data products archived together with special purpose software to perform decompression and other functions may have specialized formats. Tabular data that would be well suited to use in an off-the-shelf database product or spreadsheet package should be stored in easily interpreted data formats.

See the *Record Formats* chapter of the PDS Standards Reference. Detailed data formats are described in PDS labels after determination of the data objects that make up a data product.

### 7.2.4 Determine the Data Objects

This step will involve determining the data objects in the data products identified. PDS has designed a set of standard data object descriptions to be used to define both the contents and structure of data products. Appendix A of the PDS Standards Reference describes these standard data objects in detail along with examples of their use. Data objects are considered to be *primary* or *secondary* objects within the data product.

Examples of the types of *primary* PDS data objects and their use include:

- TABLE - a uniform collection of ROWS and COLUMNs stored in either ASCII or binary format. ASCII forms are easily imported into a variety of spreadsheets and databases.

- IMAGE - a two dimensional array of spatially organized measurements (LINES and SAMPLES). Many public domain image display programs can read PDS Image objects.

- SERIES - a special form of TABLE for storing regularly sampled measurements.

- TEXT - human readable text file in ASCII format. Compatable with all standard text editiors and word processors.

Examples of the types of *secondary* PDS data objects and their use include:

- PALETTE - a type of TABLE that contains entries representing color assignments for sample values contained in an IMAGE.

- HISTOGRAM - an array of values used for data that has been ‘binned’ often to provide a feel for certain overall qualities of a primary data object.

## 7.2.5 Determine Data Product File Configurations

Once the data objects of a given data product have been identified, decide whether primary and secondary data objects are to be in the same file or different files. Quite often, there is a one-to-one correspondence between a data product and a file. This is not required however.

For example, for the Magellan Full-Resolution Mosaic Image Data Records data set, the data product is a single mosaic composed of 56 image framelets. Each framelet is stored in a separate file and described by an IMAGE object. The 56 framelets are related to one another in a left to right, top to bottom sequence and numbered 1 through 56. A secondary data object, a histogram, is related to the full mosaic, but is stored together with each framelet file.

## 7.2.6 Design Data Product Labels

One of the most important steps in designing the data products for a data set, is the design of the data product labels. The *Data Product Labels* chapter of the PDS Standards Reference provides detailed information on this step, and Appendix A of the PDS Standards Reference provides many examples of data product labels. Figure 7.1 below is an example of a PDS label for an Image data product.

---

```
CCSD3ZF0000100000001NJPL3IF0PDSX00000001
PDS_VERSION_ID       = PDS3

/*  FILE CHARACTERISTICS */

RECORD_TYPE          = FIXED_LENGTH
RECORD_BYTES         = 956
FILE_RECORDS         = 965
LABEL_RECORDS        = 3

/*  POINTERS TO START RECORDS OF OBJECTS IN FILE */

^IMAGE_HISTOGRAM     = 4
^IMAGE               = 6

/*  IMAGE DESCRIPTION */

DATA_SET_ID          = "VO1/VO2-M-VIS-5-DIM-V1.0"
PRODUCT_ID           = "MG15N022-GRN-666A"
SPACECRAFT_NAME      = VIKING_ORBITER_1
TARGET_NAME          = MARS
IMAGE_TIME           = 1978-01-14T02:00:00
START_TIME= 1978-01-14T02:00:00
STOP_TIME= 1978-01-14T02:00:00
SPACECRAFT_CLOCK_START_COUNT= UNK
SPACECRAFT_CLOCK_STOP_COUNT= UNK
PRODUCT_CREATION_TIME= 1994-01-01T00:00:00
ORBIT_NUMBER         = 666
FILTER_NAME          = GREEN
IMAGE_ID             = "MG15N022-GRN-666A"
INSTRUMENT_NAME      = {VISUAL_IMAGING_SUBSYSTEM_CAMERA_A,
                        VISUAL_IMAGING_SUBSYSTEM_CAMERA_B}
NOTE                 = "MARS MULTI-SPECTRAL MDIM SERIES"
```

```
GEOMETRY_SOURCE_IMAGE_ID = "666A36"
EMISSION_ANGLE      = 21.794
INCIDENCE_ANGLE     = 66.443
PHASE_ANGLE         = 46.111

/*  DESCRIPTION OF OBJECTS CONTAINED IN FILE */

OBJECT              = IMAGE_HISTOGRAM
ITEMS               = 256
DATA_TYPE           = VAX_INTEGER
ITEM_BYTES           = 4
END_OBJECT          = IMAGE_HISTOGRAM

OBJECT              = IMAGE
LINES               = 960
LINE_SAMPLES        = 956
SAMPLE_TYPE         = UNSIGNED_INTEGER
SAMPLE_BITS         = 8
SAMPLE_BIT_MASK          = 2#11111111#
CHECKSUM             = 65718982
SCALING_FACTOR      = 0.001000     /* I/F = scaling factor * DN + offset, */
                                    /* convert to intensity/flux. */
OFFSET              = 0.0
STRETCHED_FLAG      = FALSE /* Optimum color stretch for display*/
STRETCH_MINIMUM = ( 53,  0)/* of color images. */
STRETCH_MAXIMUM = (133,255)
END_OBJECT          = IMAGE

END
```

Figure 7.1 Sample Data Product Label

The design of the data product labels consists of the following steps:

- Choose detached, attached, or combined detached label format - The label in Figure 7.1 happens to be attached to the data product. Some flight projects may have already determined this for you. For example, archive products that pass through the AMMOS Project Data Base (PDB) may be required to have attached labels.

- Determine label standards identifiers and Standard Formatted Data Unit (SFDU) use -These are the first two lines in Figure 7.1. Once again, there may be project specific requirements for SFDU usage and versions of PDS standards.

- Define file characteristics data elements - These are the four lines under the /* FILE CHARACTERISTICS */ comment line, which provide important physical characteristics of the file.

- Define data object pointers - These are the lines preceeded by the ^. They allow the proper starting location of each major data object in the file.

- Select identification and description data elements - These are lines under the /* IMAGE DESCRIPTION */ comment. PDS requires a few of these, such as the DATA_SET_ID and PRODUCT_ID lines. There may be additional required keywords established by your project. Additional lines may be added as desired to provide important information that should stay with the data product. You may wish to develop some new data elements.

- include PDS required data elements
- include any project required data elements
- select additional data elements from the PSDD
- work with the PDS Central Node Data Engineer to propose new data elements for addition to the PSDD or to propose modifications to existing data elements (e.g. modifying definitions, adding new values, etc.)

• Design data object descriptions - These are the lines beginning with OBJECT and ending with END_OBJECT , and provide specific information on both the structure and content of each data object in the file including the data type of the individual elements. In this case, each pixel, or SAMPLE, is a 4 byte integer in VAX format. This information is required to properly interpret the samples on various platforms.

### 7.2.7 Validate Label Syntax

Using a text editor, create a sample data product label as designed. The PDS Label Verifier can be used to ensure that your data product label follows the correct syntax. If you have already obtained and installed this PDS tool, you may do this yourself, or you may send it to the Data Engineer or Discipline Node working with you.

### 7.2.8 Validate Data Object Definitions

If the data contains TABLE, SERIES, or SPECTRUM objects, you can create a sample labeled data product and use the PDS Table Browser to validate that the label correctly defines the sample data. If the data uses the IMAGE object, create a sample labeled data product and use IMDISP, IMAGE4PDS or other image display software available through your PDS Data Engineer to display the sample image.

This is an important procedure, and many errors in both label and data content are often caught during this step. Once again, your PDS contacts may do this for you.

### 7.2.9 Write Data Product SISs

A Data Product Software Interface Specification (SIS) is a document that describes the format and size of the individual data products. This document will include the PDS data product label design that will be used. The PDS Mission Products Manager will have a signature line on this document.

Usually there is a project requirement to produce these SISs. If so, obtain the Data Product SIS outline or an example from the project. A sample Data Product SIS may also be obtained from the PDS Central Node Data Engineer. Draft Data Product SISs should be provided to the PDS Central Node Data Engineer and the appropriate Discipline Node(s) for review.

## 7.3 Design the Data Set or Collection

### 7.3.1 Define the Purpose and Scope of the Data Set or Collection

Defining the objective and scope of a data set or a data set collection is often done before, or at the same time as the definition of the individual data products. Some questions to considering in determining this are:

- What do the data describe or measure?
- What form(s) of the data should be archived (e.g. raw, calibrated, sub-sampled)?
- What range of data will be included (e.g. a single planetary encounter, data from several instruments or spacecraft of the same target, a particular time range)?

Data sets may be grouped together with other data sets into data set collections. Data set collections consist of data sets that are related by observation type, discipline, target, or time, and should be treated as a unit, to be archived and distributed together for a specific scientific objective or analysis. This is usually done during a restoration task.

An example of a data set collection is the Pre-Magellan CD-ROM containing a collection of selected Earth-based radar data of Venus, the Moon, Mercury, and Mars, Pioneer Venus radar data, airborne radar images of Earth, and line of sight acceleration data derived from tracking the Pioneer Venus Orbiter and Viking Orbiter 2.

### 7.3.2 Determine Other Data Set or Collection Components

Determine the ancillary data (e.g., navigation data, calibration data, pointing information), software, and documentation that will be included with the data set or data set collection. Required components of a Data Set/Data Set Collection are documented in the *Volume Organization and Naming* chapter of the PDS Standards Reference.

In certain cases, ancillary data may be archived as separate data sets, particularly if it is applicable to a wide variety of data sets. An example of this is a complete data set of SPICE kernels for a flight project to be utilized with a number of data sets archived for various science instruments. However, instrument specific kernel files, or those produced as special products by individual instrument teams, would be archived as part of the instrument specific data set.

### 7.3.3 Create Data Set or Collection Names and Identifiers

Choosing a name and and identifier for your data set or collection is usually straightforward. Since these values are a required part of the data product label, these are usually determined early in the design process. In the previous Image data product label example (Figure 7.1), the DATA_SET_ID = “V01/V02-M-VIS-5-DIM-V1.0”, and identifies the spacecrafts, target, instrument, processing level, product type, and version of this particular data set. See the *Data Set/Data Set Collection Contents and Naming* chapter in the PDS Standards Reference for the details on forming these names and identifiers.

## 7.4 Determine the Storage Medium

Media used for storage of archival data sets should be utilized efficiently. This includes full recording of individual media volumes, selection of recording formats to minimize wasted space on the media, and the use of simple data compression techniques to reduce the volume of infrequently accessed data. If a medium other than magnetic tape or CD-ROM is to be used, then PDS should be notified to determine whether it can be accommodated. The use of replicable media, such as CD-ROM, for products that are expected to be widely used, is recommended. For active flight projects, the medium selected for use will usually be determined by the project.

For additional information, see the *Media Formats for Data Submission and Archive* chapter of the PDS Standards Reference.

## 7.5 Design Volume /Volume Set

A volume represents a physical unit of data such as a magnetic tape, floppy disk, or CD-ROM. The hierarchical organization, as described in the *Volume Organization and Naming* chapter of the PDS Standards Reference, should be used on the storage medium if possible, e.g., CD-ROM. When using a serial medium, such as magnetic tape, a hierarchical organization can not be physically implemented. It is described, however, in the VOLDESC.CAT file included on the medium, so that the structure can be recreated if the volume is transferred from tape to disk. For the purposes of this discussion, we will consider a data volume that supports a hierarchical directory structure. For most flight projects, a Volume Organization SIS is written to document the Volume/Volume Set design.

### 7.5.1 Map Data Sets or Collections to Volumes

Taking into account the total volume of data to be archived (including ancillary data), estimates of individual file sizes, schedule of availability of the data (including any proprietary period), and operational constraints of the volume assembler, determine the allocation of data to physical volumes. This mapping is flexible:

- Data sets or collections may reside on one or more physical volumes.
- Multiple data sets may also be stored on a single volume.

When a data set spans multiple volumes, it is recommended that each volume provide a complete set of ancillary data pertaining to the data contained on the volume.

### 7.5.2 Name the Volume Set

If the data set or data set collection spans more than one volume, a name and volume set identifier must be selected to uniquely identify it. This is especially important if the volume set needs to be ordered and distributed as a single unit. See the *Volume Set Naming* chapter of the Standards Reference for further information. The PDS Data Engineer can help you with determining an appropriate naming convention for your volume set.

### 7.5.3 Name the Volumes

Each individual volume will also require a unique name and identifier. If the volume is part of a volume set, there is also a relationship between these identifiers. See the *Volume Name* chapter of the PDS Standards Reference for further information. The PDS Data Engineer can also help you determine an appropriate name for your volume.

### 7.5.4 Determine non-data Subdirectories and Files for each Volume

A number of additional directories and/or files are added to each archive volume to provide needed documentation, indices, software, etc. to allow proper use of the data volume. Some of these directories and files are required by PDS, some are recommended, and some have specific applications. The standards for PDS required and optional non-data subdirectories and files are discussed in the *Volume Organization and Naming* chapter of the PDS Standards Reference.

### 7.5.5 Determine the Data Organization

Scientific data sets are generally organized by event time, by the target of the observation, or by some cyclic value such as orbit number. The effort put into organizing the data can make the difference between simple access by the user or tedious searching. It is important to minimize the number of directory levels that must be traversed to get to the data and to minimize the amount of directory changing required during normal data access operations. For example, if two related data files are always processed in conjunction with each other, the files should be grouped in the same directory.

Organization in continuous ranges of event time is advised. An exception to this are data that clearly relate to some well-understood observation, target, or event. If the data are organized on some other basis, the preparer should make available a mechanism to extract data based on event time in order to support correlative analysis.

- Time-Oriented Organization

The simplest approach is to create a single data file containing a fixed time interval, such as a day, week, or month (Figure 7.2)

.

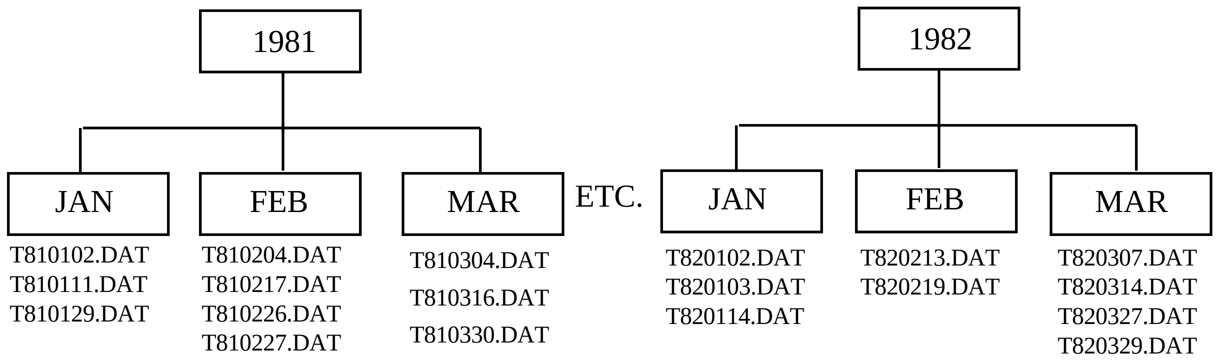



Figure 7.2 Example of Time-Oriented Organization

- Target-Oriented Organization

In the target-oriented approach, the target of the instrument observations is used to build a directory structure, and all observations of that target within a certain time span of data are included (Figure 7.3).

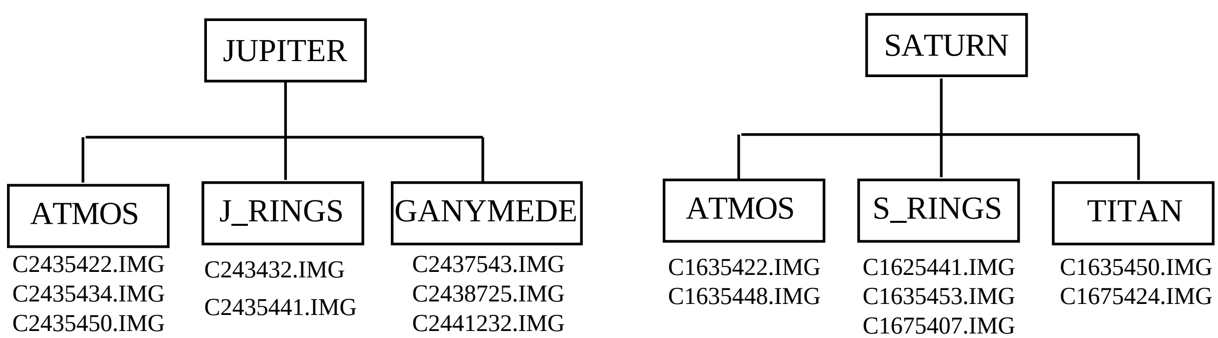



Figure 7.3 Example of Target-Oriented Organization

- Event-Oriented Organization

Event-oriented organization collects data based on events related to spacecraft operations or observed phenomenon (Figure 7.4).

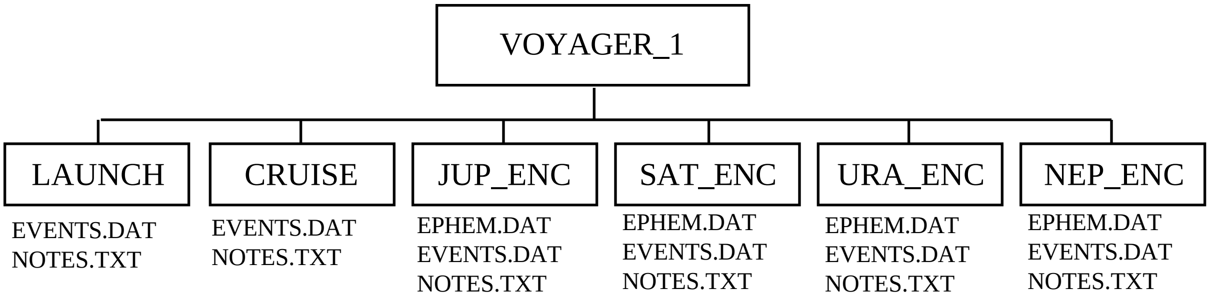



Figure 7.4 Example of Event-Oriented Organization

- Bin-Oriented Organization

Bin-oriented schemes collect data in spatially defined bins, such as 5-degree lat-lon bins (numbered sequentially). This is similar to target-oriented organization, except that bins are used to represent the observation targets. If arbitrary bin names are used (BIN1, BIN2, etc.), a mapping between bins and spatial coverage must be supplied. It is often possible to include the limits of the binning parameter in the name (Figure 7.5), where longitude is used to group and name files.

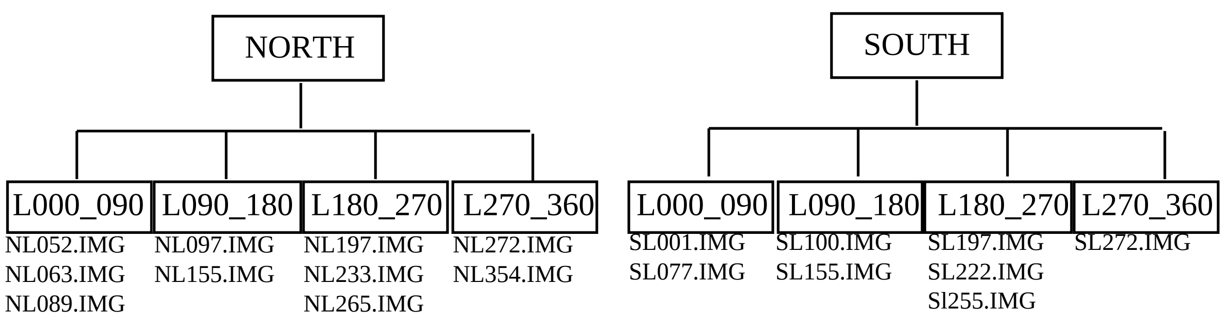



Figure 7.5 Example of Bin-Oriented Organization

### 7.5.6 Establish Data Product File Naming Convention

The file naming scheme utilized should result in file names that uniquely identify the data included in the file but have some characteristics in common with other files. For example, T810111.dat contains data from January 11, 1981 and allows use of file selection capabilities like the use of wildcards (T81011*.dat selects January 10th through 19th).

Special consideration should be given to constructing file names which contain dates. File names which begin with the names or partial names of months (such as JAN1989.DAT) can cause problems when outside applications process a series of files alphabetically. A series of files with file names starting with month names will be arranged with APRIL first, then AUGUST and so on, rather than the desired JANUARY, FEBRUARY, etc.

If possible, all file names on a volume should be unique, even if the files are stored in different subdirectories. The reason for this is to avoid confusion and possible data loss if files are transferred from a hierarchical directory structure to a flat one.

If detached labels are used, it is recommended that the name of the label files be the same as the data object files, except with a .LBL extension.

See the *File Specification and Naming* chapter of the PDS Standards Reference for further information.

### 7.5.7 Establish Directory Naming Convention

Directories should contain about a “screenful” of file entries. If too many files are included in a directory, it becomes difficult for users to “browse” through the file names using the directory command. If too few files are included, this will severely increase the complexity of locating a file. As a rule, the depth of directories should be kept to a minimum unless the number of files per directory becomes too large (about 100 is a reasonable maximum).

Balance simplicity with explicitness in attempting to encode a description of the directory contents in the name. It is often better to use simple arbitrary names, than complex but meaningful ones (PWS1, PWS2, PWS3 vs. VGPWSXZ1, VGPWSXZ2, VGPWSXZ3).

See the *Directory Types and Naming* chapter of the PDS Standards Reference.

### 7.5.8 Determine the Indices Needed

An index is an important part of a PDS archive volume. A good index or a set of indices provides the user with the means to rapidly locate specific data files of interest. For instance, columns containing latitude and longitude ranges or time periods followed by a column of directory and file names will provide the user with information needed to find individual data file of interest for a particular investigation. Wherever possible, indices should be built directly from the data product labels.

PDS indices are ASCII tables, and are required to be labeled. For data sets that span multiple volumes, it is important to design the index so that the same table format may be used across all the volumes in the set. Another type of index used on volume sets is a cumulative index. For instance, if the current volume is only one of a volume set, the cumulative index can provide context for the user as to how this volume fits into the set and what volumes have been produced previously.

Additional information on creating indices and examples can be obtained from the Central Node Data Engineer.

Once you’ve designed your indices, create a test index table with several sample index records, and it's corresponding index label. You can now use the PDS Label Verifier and the PDS Table Browser to verify that these are correct. The PDS Data Engineer can help you in this validation process.

### 7.5.9 Write Volume Organization Software Interface Specifications (SISs)

Volume Organization SIS describes the design of an archive volume or volume set. In Magellan, these SISs were called the CD-ROM SISs. In Mars Observer they were called the Archive Collection SISs. The PDS Mission Products Manager will have a signature line on these documents.

Usually there is a project requirement to produce these SISs. If so, obtain the Volume Organization SIS outline or an example from the project. A sample SIS may also be obtained from the PDS Central Node Data Engineer. Draft Volume Organization SISs should be provided to the PDS Central Node Data Engineer and the appropriate Discipline Node(s) for review.

## 7.6 Design Data Production Process

### 7.6.1 Design Label Generation Process

For data restorations, label generation is often performed by the PDS Discipline Node or subnode. For flight projects the PDS Discipline Node or Central Node Data Engineer can offer suggestions for automating the process of generating labels, including the use of PDS tools (see Appendix B). Specialized label generation may benefit from using the PDS Label Library. Other PDS software utilities available specifically for label generation are:

- The PDS Table to Label Generator (TAB2LAB) may sometimes be used to produce labels from a table of ASCII data, typically an index table.
- The PDS File to Label Generator (FILE2LAB) may be used to produce labels from existing ASCII data files.

### 7.6.2 Design the Volume Production Process

For data restorations, volume production is typically performed by the PDS Discipline Node. For flight projects, the PDS Discipline Node or Central Node Data Engineer can offer suggestions and software tools to help automate the process of generating volumes.

## 7.7 Plan Data Validation Process

For active flight projects, data validation plans are usually documented in project operational procedures. For data restorations, planning data validation is part of the peer review process which sets the criteria for accepting data sets.

### 7.7.1 Investigate Validation Tools

PDS has developed several validation tools that have been useful in a variety of validation steps. Some of the currently available programs (see Appendix B) and their use include:

PDS Label Verifier - This utility verifies the syntax of a PDS label, including labels attached to their data file, insuring the *language* of the label is correct. In addition, an important feature of this utility is a link to the PSDD, to ensure that the *values* provided in the label are of the proper type and range. Many errors are it the information being provided are caught at this stage of validation, and can make the difference between a data product that is accurate, and one that will ultimately confuse the end user.

PDS Table Browser - This utility provides a way to interactively examine both ASCII and binary tabular data (defined as TABLE, SERIES, or SPECTRUM data objects) through the PDS label that describes the data format. It provides a good way to view test data files (binary files in particular) to make sure the COLUMNs are properly defined, the record types and lengths are correct, etc. Use of the *summary* feature in this program may also help you find errors in the ranges of some of your data, as well as fields that may be missing data.

PDS Volume Verifier - This utility is currently under development and will perform verification of a complete CD-ROM or test CD-WO volume. Contact your Central Node Data Engineer for further information.

### 7.7.2 Write Data Validation Plan

For flight projects, a Data Validation Plan is a document that defines the operational procedures that will be in place for validating the content and physical organization of the volumes produced. The PDS Mission Products Manager will have a signature line on this document.

Usually there is a project requirement to produce a Quality Control or Validation Plan. If so, obtain an outline or example from the project. A sample Validation Plan may also be obtained from the PDS Central Node Data Engineer. Draft Data Validation Plans should be provided to the PDS Central Node Data Engineer and the appropriate Discipline Node(s) for review.

When planning data validation, the following should be addressed:

- Definition of the validation process, including the establishment of both automatic and manual validation procedures
- Assignment of roles and responsibilities for each validation step

### 7.7.3 Design, Implement, and Test Project Data Validation Software

Typically, data validation software is tailored to the particular needs of the project. This software will need to be designed, implemented, and tested by the project.

### 7.7.4 Plan for updates to Data Validation Procedures and Software

Planning data validation is often an iterative process. Both the procedures and software used to perform the validation will probably require a few updates.

In the case of Magellan, test volumes and the cooresponding validation procedures evolved until both the data producers and users were convinced that an error-free volume had been produced. This involved the passing a final comprehensive quality-control check.

## 7.8 Complete the Catalog Object Templates

All data submitted to the PDS is accompanied by a set of *catalog object templates*, or forms, used for loading the PDS Data Set Catalog. By filling out these templates, the data supplier provides high-level information concerning the mission, instrument host (i.e. spacecraft of earth based instrument host), instruments, and project personnel, as well as information about each of the data sets or data set collections being submitted to PDS. Figure 7.6 is an example of a completed catalog object template for a data set to be archived.

Examples and assistance in completing these will be provided by the PDS Data Engineer or the PDS Discipline Node with which you are working. In some data restoration cases, a Discipline Node may simply ask you for descriptive information in paragraph format, and they will put it into the proper template format. Also, automated tools may be available to support the editing and validation of your templates. Contact your PDS Data Engineer for more information.

```
/* Template: Data Set Template                              Rev: 19890121 */

OBJECT                        = DATA_SET
DATA_SET_ID                   = "VG2-U-PWS-2-EDR-WFRM-60MS-V1.0"

OBJECT                        = DATA_SET_INFORMATION
DATA_SET_NAME                 = "VG2 URA PWS RAW EXPERIMENT WAVEFORM 60MS V1.0"
START_TIME                    = 1985-11-08T07:04:08.649
STOP_TIME                     = 1986-02-22T20:41:34.186
DATA_OBJECT_TYPE              = "TIME SERIES"
DATA_SET_RELEASE_DATE         = 1992-04-06
PRODUCER_FULL_NAME            = "DR. WILLIAM S. KURTH"
DETAILED_CATALOG_FLAG         = N
DATA_SET_COLLECTION_MEMBER_FLG = N
DATA_SET_DESC                 = "

Data Set Overview
=============
This data set consists of electric field waveform samples from the Voyager 2 Plasma Wave Receiver waveform receiver obtained
during the Uranus encounter. The waveforms are collections of 4-bit samples of the electric field measured by the dipole electric
antenna at a rate of 28,800 samples per second.  1600 samples are collected in 55.56 msec followed by a 4.44-msec gap. Each 60-
msec interval constitutes a line of waveform samples. The data set includes about 271 frames of waveform samples consisting of
up to 800 lines, each.  The telemetry format for the waveform data is identical to that for images, hence the use of line and frame
as constructs in describing the form of the data.  The waveform is sampled through a bandpass filter with a passband of 40 Hz to
12 kHz.  The 4-bit samples provide sixteen digital values of the electric field with a linear amplitude scale, but the amplitude scale
is arbitrary because of the automatic gain control used in the waveform receiver.  The instantaeous dynamic range afforded by the
4 bit samples is about 23 db, but the automatic gain control allows the dominant signal in the passband to be set at the optimum
level to fit within the instantaneous dynamic range.  With the gain control, the overall dynamic range of the waveform receiver is
about 100 db.  The automatic gain control gain setting is not returned to the ground, hence, there is no absolute calibration for the
```

data. However, by comparing the waveform spectrum derived by Fourier transforming the waveform to the spectrum provided by the spectrum analyzer data, an absolute calibration may be obtained in most cases. The data may be plotted in raw form to show the actual waveform; this is useful for studying events such as dust impacts on the spacecraft. But the normal method of analyzing the waveform data is by Fourier transforming the samples from each line to arrive at an amplitude versus frequency spectrum. By stacking the spectra side-by-side in time order, a frequency- time spectrogram can be produced. Additional information about this dataset and the instrument which produced it can be found elsewhere in this catalog. An overview of the data in this data set can be found in Gurnett et al. [1986] and a complete instrument description can be found in Scarf and Gurnett [1977].

Parameters
========

| | |
|---|---|
| Sampling Parameter Name | : TIME |
| Sampling Parameter Resolution | : 0.000034722 |
| Minimum Sampling Parameter | : N/A" |
| Maximum Sampling Parameter | : N/A" |
| Sampling Parameter Interval | : 0.000034722 |
| Minimum Available Sampling Int | : 0.000034722 |
| Sampling Parameter Unit | : SECOND |
| Data Set Parameter Name | : PLASMA WAVE WAVEFORM |
| Noise Level | : 0.000005 |
| Data Set Parameter Unit | : VOLT/METER (Data not absolutely calibrated)) |

“

CONFIDENCE_LEVEL_NOTE= "

Confidence Level Overview
====================
This data set includes all available waveform receiver data obtained during the Uranus encounter. There has been no attempt to clean various interference signals from the data. Most of these can normally be easily seen in frequency-time spectrograms as narrowband, fixed-frequency tones. The most common include narrow-band tones at 2.4 and 4.8 kHz which are power supply harmonics. There is sometimes a tone near 1.7 kHz which is associated with the operation of the spacecraft gyros. The spacecraft tape recorder results in a rather intense band in the frequency range of a few hundred Hz. There are few times when the data in this frequency range can be used. However, there are times when the real signals in this frequency range can exceed the intensity of the interference sufficiently so that the frequency range near a few hundred Hz can be used. Use of the spectrum analyzer data can be of use to determine when these time periods occur. The stepper motor of the LECP instrument also interferes in the frequency range of a few hundred Hz, but for periods of a few seconds. The LECP interference is very intense and captures the automatic gain control so that real signals, even where there is no interference, will appear to decrease in amplitude until the LECP interference fades in amplitude. The PLS instrument periodically interferes at 400 Hz and odd harmonics because of a 400-Hz square wave used to modulate a grid in the detector. The PLS interference lasts for several seconds and ends abruptly. Telemetry errors result in a fairly graceful degradation of the waveform data. Assuming the telemetry errors are randomly occurring bursts, they typically appear as an enhanced background level in the spectrum. Since the bursts are short, their Fourier transform is a broadband spectrum. When looking for relatively narrowband features or features with distinct frequency-time characteristics, the result of the bursts simply reduce the signal-to-noise in the spectrum. One way of reducing the effect of burst telemetry errors is to pass the waveform data through a low-pass filter to despike it, prior to running the Fourier transform. The waveform data is not subject to the negative effects of the failure in the Voyager 2 Flight Data System which reduces the sensitivity of the spectrum analyzer and affects the calibration above 1 kHz. In fact, use of the 1 - 12 kHz waveform data is an effective way of avoiding the problems with the spectrum analyzer data in this frequency range."

```
END_OBJECT                = DATA_SET_INFORMATION

OBJECT                    = DATA_SET_TARGET
TARGET_NAME               = URANUS
END_OBJECT                = DATA_SET_TARGET

OBJECT                    = DATA_SET_HOST
INSTRUMENT_HOST_ID        = VG2
INSTRUMENT_ID             = PWS
END_OBJECT                = DATA_SET_HOST

OBJECT                    = DATA_SET_REFERENCE_INFORMATION
```

```
REFERENCE_KEY_ID            = GURNETTETAL1986
END_OBJECT                  = DATA_SET_REFERENCE_INFORMATION

END_OBJECT                  = DATA_SET
```

Figure 7.6 Example Catalog Object Template for a Data Set

### 7.8.1 Obtain the Catalog Object Templates

You will receive a set of blank catalog templates from your PDS Central Node Data Engineer in the form of one or more ASCII files. These are usually sent to you electronically, so that they may be easily edited. Only those templates that need to be completed for the data set or sets you are submitting will be sent to you. Some templates may need to be duplicated several times (e.g., three Data Set templates because you know you are submitting three data sets). Templates may also require duplication of objects within them (e.g., multiple References for a Mission description).

Any standard text editor can be used to complete these catalog object templates. Further descriptions and examples of these objects are provided in Appendix B of the PDS Standards Reference, or may be obtained from your PDS Data Engineer.

### 7.8.2 Review the structure of the Catalog Object Templates

Take a moment to review the layout of the templates before you edit them. Each object template begins with a set of comment fields which describe the object (or set of objects).These comment fields contain the template name, notes about the object, and a structure of the object, if needed. Many objects have a hierarchical structure, and there may be multiple levels of nesting.

The body of each object is delimited by the OBJECT and END_OBJECT data elements. The OBJECT data element indicates the beginning of a new object and should not be modified. The END_OBJECT data element indicates the end of a single object or a nested set of objects and should not be modified.

### 7.8.3 Complete the Catalog Object Templates

Using any standard text editor, fill out the templates as with as much infomation as you can provide. Comment fields are included in the blank templates at the right side of each line, and are delimited by /* */. These comments provide information on the types of values that should be supplied for each line, and may be deleted.

Your PDS Data Engineer can assist you with any questions you may have about filling in values, and may refer you to the Planetary Science Data Dictionary (PSDD) for further information. The PSDD provides the definitions of the objects, definitions of each data element in the templates, and the allowable values (standard values) for many of these data elements.

In certain cases, a data element may not be applicable within the context of the data being submitted. In such cases, the value “N/A” is entered. In other cases, where they value is not known, “UNK” may be used. See the *Usage of N/A, UNK, and NULL* chapter of the PDS Standards Reference.

### 7.8.4 Validate the Completed Catalog Objects

The completed catalog objects can be verified using the PDS Label Verifier (see Appendix B) to ensure the syntax and values supplied are valid. This step will often identify many new standard values that will need to be added to the PSDD in support of your new data. The PDS Data Engineer can assist you with this step, or may do it for you.

### 7.8.5 Submit the Completed Catalog Objects

The completed catalog objects should be returned to the PDS Central Node Data Engineer as an ASCII file. Again, this is transfer is typically done electronically. The comment area of the object templates can be used to include a date or version number. The PDS Data Engineer will review the completed catalog object templates and return comments to you. Preparing the catalog information is an iterative process. The catalog information may be submitted and reviewed several times until both the data supplier and the PDS Data Engineer are satisfied with their correctness and completeness.

### 7.8.6 Plan for updates to the Templates.

For a mission in progress, more information may become available and the catalog object templates may need to be updated. If so, contact your PDS Data Engineer.

# Chapter 8

## Data Set Assembly and Validation Checklist

- ☐ 8.1 Create the Labeled Data Products
- ☐ 8.2 Create a Data Staging Area for Volume Production
- ☐ 8.3 Write Volume Documentation
- ☐ 8.4 Collect Catalog Files
- ☐ 8.5 Collect and Format Documentation
- ☐ 8.6 Prepare Supporting Software
- ☐ 8.7 Collect Geometry and Calibration Files
- ☐ 8.8 Generate the Data Indices
- ☐ 8.9 Validate Non-product Labels
- ☐ 8.10 Prepare a Set of Test Volumes and Distribute
- ☐ 8.11 Execute Data Validation Procedures
- ☐ 8.12 Transfer the Data to the Final Medium

# Chapter 8

# Data Set Assembly and Validation

Data set assembly consists of collecting and formatting all of the components of the data set, processing it according to the design, and storing it on the planned medium. It also involves preparing PDS data product labels, writing the volume documentation, and creating the volume indices.

## 8.1 Create the Labeled Data Products

Process and format the data objects into products as designed. Processing may involve steps to calibrate the data, or fill in missing data points, for example. Another step may be to create a histogram of one of the data objects. These objects may then all be packaged together into the final data product, including the PDS label. Creation of the labels may be concurrent with processing the data objects themselves, or may be done independently, and attached (if attached labels are used) after other processing has been completed.

## 8.2 Create a Data Staging Area for Volume Production

Before volume production begins, a data staging area should be created. This area should provide a structure similar to that which will be used on the final medium; this allows the volume assembler to store the pieces of the data set as they are assembled.

## 8.3 Write Volume Documentation

- Prepare a VOLDESC.CAT file to summarize the contents of each volume submitted. (See the VOLUME Object Definition in the PDS Standards Reference, Appendix A.) This file will describe the contents of the volume. This file will be the first one written to the volume and will contain physical volume data and the structure of directories and files. (This is especially important for magnetic tape volumes.)
- Prepare the AAREADME.TXT file. This file is a high-level description of the contents of the volume in plain ASCII text. See the PDS Standards Reference, Appendix D an outline and example of this file.
- Prepare a VOLINFO.TXT file (if this file has been identified for inclusion on the volume during the Design process). The VOLINFO.TXT file provides a detailed summary of the contents of each volume in plain ASCII text. An outline or example can be obtained from the PDS Central Node Data Engineer.
- Prepare other text files (such as ERRATA.TXT) that have been identified for inclusion on the volume.

## 8.4 Collect Catalog Object Files

- If catalog object files have been identified for inclusion in the volume during the design process, these files need to be assembled. The final catalog object templates will need to be obtained from the PDS Central Node Data Engineer. These will have been corrected and are the ones you should use in your data set.

- Prepare a CATINFO.TXT file to describe the contents of the CATALOG subdirectory.

## 8.5 Collect and Format Documentation

- If document files have been identified for inclusion on the volume during the design process, these files need to be assembled and formatted. An ASCII version of all document files should be prepared.
- Prepare a DOCINFO.TXT file to describe the contents of the DOCUMENT subdirectory.

See the *Documentation* chapter of the PDS Standards Reference for further information.

## 8.6 Prepare Supporting Software

- You may want to include software for displaying, subsetting, or analyzing your data set on the volume. While this is not required, it often makes the difference between a data set that is actually *used* by others, and one that sits on the shelf.
- Software submitted along with your data set should include the source code, and batch files or instructions for building executable programs on applicable computer systems. Don't forget to include little procedures ( e.g. UNIX scripts, VMS command procedures) that you have found useful with the data set. Comments or guides which might help a user who is unfamiliar with the software are also recommended.

Software documentation such as requirements, design, and user manuals, where available, can also be provided.

- Prepare a SOFTINFO.TXT file to describe the contents of the SOFTWARE directory.

## 8.7 Collect Geometry and Calibration Files

- If data set or volume specific calibration files have been identified for inclusion on the volume, these need to be collected. These may take the form of PDS labeled data products , such as a TABLE, or may be textual calibration reports, often available in the form of internal memoranda. The form is not as important as the content. What PDS is striving for is a well documented data set.
- Prepare a CALINFO.TXT file to describe the contents of the CALIB subdirectory.
- If data set of volume specific geometry files have been identified for inclusion on the volume, these also need to be collected. Geometry data may include spacecraft and target ephemeris data, instrument geometry information, platform pointing information, and events files. Some flight projects make use of special ancillary information system products (e.g. SPICE kernels, SEDRs).
- If your data set is part of an archive for a large flight project, these files may already be part of

a planned archive of *ancillary-only* data and software. Sometimes having a particular subset of this data is more conveniently located on the same volume as the data set. If this is the case, be sure to make sure you have the most up-to-date files for inclusion on your volume. Contact the Navigation Ancillary Information Facility (NAIF) node of PDS for assistance.

- Prepare a GEOMINFO.TXT file to describe the contents of the GEOMETRY subdirectory.

## 8.8 Generate the Data Indices

- Generate the indices for your volume (INDEX.TAB) as designed. List the full path names of the data object files and associate them with their specific attribute value, such as latitude/longitude, or orbit number, for instance.
- For a volume (cumulative) index, list the volume IDs and their specific attribute values.
- Prepare a INDXINFO.TXT file to describe the contents of the INDEX subdirectory.

## 8.9 Validate Non-product Labels

The VOLDESC.CAT, INDEX.LBL, CUMINDEX.LBL, and any label files in the DOCUMENT subdirectory should be validated with the PDS Label Verifier. The index files should be checked with the PDS Table Browser.

## 8.10 Prepare a Set of Test Volumes and Distribute

Test volumes should be prepared and distributed to the PDS Central Node Data Engineer, Discipline Node(s), and project PI teams (for active flight projects) for volume validation.

## 8.11 Execute Data Validation Procedures

Volumes will be validated according to the operational procedures and validation criteria outlined in the active Flight Project's Data Validation Plan or Data Restoration validation criteria.

PDS will also validate volumes to ensure adherence to PDS standards.

## 8.12 Transfer the Data to the Final Medium

For magnetic tapes, copy your data files to ANSI labeled tapes. The ANSI standard allows the data product to be composed of meaningful files which can be extracted on various computer systems. ANSI labeled tapes can be written on a VAX/VMS computer using the COPY command. (DO NOT use the VMS backup command.) Contact your PDS Data Engineer for the appropriate commands to use on other computers.

For CD-ROMs, a pre-master tape must be prepared to be sent to a CD-ROM vendor. Contact the PDS Project Engineer (see Appendix A) if you would like more information on how to prepare data for CD-ROMs.

# Chapter 9

## Data Delivery Review Checklist

- ☐ 9.1 Establish a Review Committee
- ☐ 9.2 Prepare for the Data Delivery Review or Peer Review
- ☐ 9.3 Participate in a Data Delivery Review or Peer Review
- ☐ 9.4 Correct /Document Review Liens

# Chapter 9

# Data Delivery Review

Prior to archive by PDS, the data sets need to be reviewed. The purpose of the review process is to ensure the accuracy, dependability, and usefulness of science data to be distributed by PDS. The primary goal is to make sure the data is well documented for future users.

This review process is flexible, and depends on both the amount and complexity of the data being archived. For data restorations, PDS has established a process called a Peer Review. For active flight projects, the review process established by the project typically serves this purpose, and is referred to by PDS as a Data Delivery Review. Steps in the review process include:

- Establishing a review committee
- Preparing for the review
- Participating in the review
- Correcting or documenting liens from the review

## 9.1 Establish a Review Committee

For data restorations, the PDS Discipline Node responsible for archiving the data set(s) will establish a review committee. Members typically include the Discipline Node Manager, data preparer or Principal Investigator (PI), data supplier, scientists from the planetary community capable of critically reviewing the data, and a PDS data engineer. A chairperson of the committee will be named for coordinating the review process.

For active flight projects, a committee established by a project archive working group may often perform a substantial part of this review function. For example, on the Magellan project, the Radar Investigation Group (RADIG) reviewed Magellan CD-ROM volumes and either approved or disapproved of their release to the scientific community and for archive. Depending upon the review criteria established by the project, PDS may wish to perform an additional review of the archive volumes, with an emphasis on documentation and adherence to PDS standards. The responsible PDS Discipline Node may establish a review committee, as in a data restoration, to serve this purpose.

## 9.2 Prepare for the Data Delivery Review or Peer Review

Reviews may be handled in a variety of ways, and are usually determined by the availability and location of participants on the review committee. In some situations, a science conference or project meeting may be scheduled in which a good percentage of the review committee can be present to review the data and documentation. Most often, in order to provide reviewers enough time to adequately review the data, a review period is established. A meeting (or a teleconference) can then be held at the end of the review period to discuss results and identify any liens. Three to four weeks is a typical time period for a review, but this depends on the size or complexity of the data, and whether results from any prior reviews are available.

The chairperson of the review committee will provide instructions to the review committee. This includes providing either copies of the data set(s)(e.g. CD-WO media) or electronic access to the data (or samples of them), supporting documentation, completed high-level catalog templates, detailed-level catalog samples if applicable, PDS labels, and review critieria. If special software is available for viewing, analyzing, or ordering the data, distribution or access to this software will be provided also.

As part of the preparation for a review, completed catalog templates will be test-loaded into a review database by the PDS data engineer to ensure the associated information is complete. This test database (a test version of the PDS Data Set catalog) will also be available for review, if desired, and provides a way to see how a user would locate and/or order the data through PDS.

## 9.3 Participate in a Data Delivery Review or Peer Review

Once preparations for the review are complete, the review is conducted according to the instructions provided by the review chairperson. If the review is conducted as a meeting, data providers or their representatives may make presentations on the data set(s) which they have provided. This may include on-line demonstations of associated software to display, subset, or order the data. For data being reviewed over a longer time period, reviewers may read provided documentation, test out supplied software, and where possible, try using the data for science purposes and with locally available software.

At the end of the review meeting, or review period, the review chairperson summarizes the results of the review, and any liens identified are documented. Liens are usually classified as major and minor. One or more reviews may be held before all the data are delivered.

## 9.4 Correct/Document Review Liens

Wherever feasible, liens are corrected prior to delivery of the data set to PDS for archiving. If it is not feasible to correct all of the liens, the results of the review will be archived with the data sets to document all known errors and discrepancies. The review committee is responsible for deciding whether to go ahead and archive the data with documented liens.

Notes may also be added to the PDS Data Set Catalog to indicate the quality of data, adherence to PDS standards, or data usability.

# Chapter 10

## Archive and Distribution Checklist

- ☐ 10.1 Coordinate Generation of CD-ROM Copies with PDS (if applicable)
- ☐ 10.2 Physically Transfer Archive Volumes to PDS and NSSDC
- ☐ 10.3 Assist the PDS Discipline Node in Generating the DIF
- ☐ 10.4 Update Corrected or Enhanced Data Sets

# Chapter 10

# Archive and Distribution

After the Data Delivery Review or Peer Review is complete, data sets are delivered to PDS for archiving. The data sets are then made available for distribution to the planetary science community, unless other arrangements have been made. For example, very low-level data sets may be transferred to PDS for archiving only, rather than for general distribution.

## 10.1 Coordinate Generation of CD-ROM Copies with PDS (if applicable)

If the flight project chooses to use CD-ROMs for internal distribution, PDS and the NSSDC will provide funding for additional copies to be made for the general science community. The project should keep PDS informed about production plans in order to avoid multiple setup costs for additional CD-ROM copies.

In some cases, the project and PDS can provide distribution lists to the CD-ROM vendor to facilitate early distribution.

## 10.2 GTDA Classification Procedures

The following procedures are required to meet the PDS goals of wide distribution and at the same time ensure that PDS products meet Federal data distribution requirements of the Departments of Commerce, State, and Defense.

By working with the JPL International Affairs Office, PDS has secured the General Technical Data Available (GTDA) classification for all current PDS data holdings, software, and on-line systems. The significance of the GTDA classification is that PDS data and software may be exported without securing an export license for each distribution.

With certain exceptions, the GTDA classification is a blanket for all current and future PDS data, software, and systems. The exceptions are data that describe the construction of instruments and spacecraft. PDS will provide information on review requirements for this type of data.

For CD-ROMs:
Add “GTDA” to all CD-ROM artwork and CD-ROM backliners. A sample is available from your Central Node Data Engineer.

For tapes:
Add “GTDA” to all physical tape labels.

## 10.3 Physically Transfer Archive Volumes to PDS and NSSDC

Current flight projects will implement the data delivery steps specified in their Archive Policy and Data Transfer Plan, including the delivery to a PDS Discipline Node or Nodes, and possibly to NSSDC. These procedures will ensure the project notifies the PDS Operator when the physical products are shipped to a PDS node.

Restored data sets will be delivered to the appropriate PDS Discipline Node. When delivery is made to a PDS Discipline Node, the physical data set will be added to the Discipline Node’s data inventory. The node will then be responsible for archiving the data set at the NSSDC.

## 10.4 Assist the PDS Discipline Node in Generating the DIF

Information about the archive data is provided to the NSSDC in Directory Interchange Format (DIF). The DIF information is required by the NSSDC and is used by them to populate their Master Directory. The PDS Discipline Node will be responsible for generating the DIFs, but may request assistance from the data supplier.

## 10.5 Update Corrected or Enhanced Data Sets

Data sets can be updated to include corrections or enhancements after they have been archived. Contact your PDS Data Engineer or Discipline Node for information on this process.

# Appendix A

# Whom to Contact

## Mission Products Manager (and Data Engineer contacts)

Yolanda Fletcher
M/S 525-389
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109
TCP: yfletcher@jplpds.jpl.nasa.gov
Phone: 818/306-6023

## PDS Operator

PDS Operator
M/S 525-389
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109
TCP: pds_operator@jplpds.jpl.nasa.gov
Phone: 818/306-6130

## PDS Project Engineer

Mike Martin
M/S 525-389
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109
TCP: mmartin@jplpds.jpl.nasa.gov
Phone: 818/306-6038

## Atmosphere Node Manager

Reta Beebe
New Mexico State University
University of Colorado
Department of Astronomy
P.O. Box 30001/Dept 4500
Las Cruces, NM 88003
TCP: rbeebe@nmsu.edu

## Geosciences Node Manager

Ray Arvidson
Washington University
Campus Box 1169
One Brookings Drive
St. Louis, MO 63130-4899
TCP: arvidson@wunder.wustl.edu

## Imaging Node Manager

Larry Soderblom
USGS Astrogeology
2255 N. Gemini Drive
Flagstaff AZ 86001
TCP: lsoderblom@astrog.span.nasa.gov

## Navigation and Ancillary Information Facility Manager

Chuck Acton
M/S 301-125L
Jet Propulsion Laboratory
4800 Oak Grove Drive Pasadena, CA 91109
TCP: cha@naif.jpl.nasa.gov

## Planetary Plasma Interactions Node Manager

Ray Walker
UCLA
IGPP
6851 Slichter Hall
Los Angeles, CA 90024
TCP: rwalker@igpp.ucla.edu

## Planetary Rings Node Manager

Mark Showalter
NASA Ames Research Center
245-3
Moffett Field, CA 94035
TCP: showalter@ringside.arc.nasa.gov

## Small Bodies Node Manager

Michael A'Hearn
Dept of Astronomy
University of Maryland
College Park, MD 20742-2421
TCP: ma@astro.umd.edu

## Data Distribution Laboratory

Kristi Kawasaki
M/S 525-389
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109
TCP: kkawasaki@jplpds.jpl.nasa.gov
Phone: 818/306-6043

# Appendix B

# PDS Tools

## B.1 The PDS Software Inventory

The PDS Software Inventory lists software that is available within the Planetary Science community and under the auspices of the PDS. Software entities included are applications, tools, and libraries that have proven useful for the display, analysis, formatting, transformation, and preparation of either science data or metadata for the PDS archives. This inventory can be accessed over theWorld Wide Web from the following URL:

http://stardust.jpl.nasa.gov/

## B.2 PDS Toolbox Releases

PDS will provide periodic releases of the PDS toolbox software and associated documentation. Releases may include either source code (C or C++), executables or both, and are available for electronic distribution. The executables for the PDS Toolbox are generated on:

- A Sun SPARCstation using SunOS 4.1 and Sun C compiler.
- A VAX 4000 running VMS 5.4 and VMS C compiler.
- A PC-386 running MS-DOS 5.0 and Borland C(++) 2.0.

See B.2.1 for specific instructions on obtaining the following tools:

- <u>PDS Label Verifier (lvtool), Version 1.2</u>

A tool for checking the syntax and semantics of PDS labels by comparing them to the Toolbox Data Dictionary.

- <u>PDS Table Browser (tbtool), Version BETA</u>

A tool for interactively examining PDS labelled ASCII and binary data in the form of TABLEs, SERIES, or SPECTRUMs, in order to verify the correctness of the label or browse, summarize, and visually verify columns of data.

- <u>PDS Table to Label Generator (tab2lab), Version 2.1</u>

A tool for generating PDS labels given a skeleton (example) label and an ASCII table of data values for insertion into the skeleton.

- PDS File to Label Generator (file2lab), Version BETA

A tool for generating PDS labels given a skeleton (example) label and a set of ASCII data files (or file headers) that contain values for insertion into the skeleton.

- PDS Simple Label Editor (sled), Version 2.4

A command line editor for verifying, modifying, and reformatting PDS labels. Includes hooks to the PDS Label Verifier. (NOTE: SLED does not yet support PDS Version 3 SFDU labelling standards.)

- PDS Utilities

    - PDS label formatter (labform), for pretty-printing labels. (Version 1.0)
    - PDS label expander (explab), for expanding a label to include format files referenced by ^STRUCTURE lines in a label. This allows proper validation of the hierarchy for an entire label using the Label Verifier rather than just a subset of a label (Version 1.0).
    - Label Dealiasing Utility (dealias), for use in updating labels with old keywords to new standards utilizing the Toolbox Data Dictionary (Version 1.1)
    - Add Columns to Table Utility (addcols), for converting old format PDS tables (OBJECT = NAME) to new format (OBJECT = COLUMN). (Version 1.1)
    - Make Data Dictionary Index Utility (make index), for generating the index associated with the Toolbox Data Dictionary (Version 1.2).

- PDS Toolbox Data Dictionary, Version 3.0, generated from the on-line Planetary Science Data Dictionary database, this form of the PSDD is required for use with the PDS Label Verifier. The Toolbox DD will be regenerated whenever the on-line PSDD is updated.

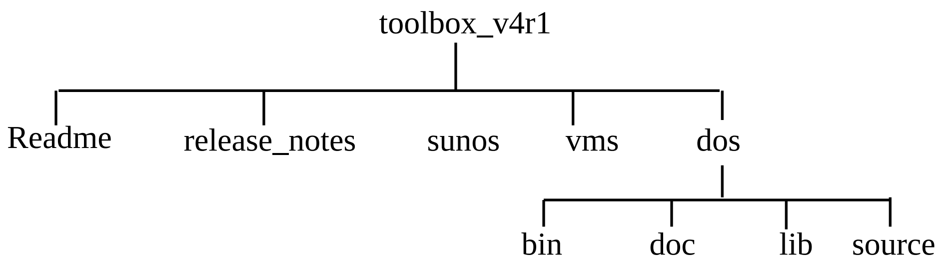
The following documents are also available with PDS Toolbox releases:

The PDS Label Verifier User's Guide
The User's Guide for the PDS Label Generators
The PDS Toolbox Overview
The User's Guide for the PDS Simple Label Editor
The PDS Table Browser User's Guide
The PDS Toolbox Utilities User's Guide
The PDS Label Library User's Guide

### B.2.1 Obtaining the Toolbox

1. Anonymous ftp to starhawk.jpl.nasa.gov (137.79.108.113)

2. Change directory to access the latest toolbox version, currently Version 4 release 1 (“cd toolbox_v4r1”). The directory structure underneath it will look like this:




Under each operating system type (sunos, vms, or dos):

The bin subdirectory contains the binary executables for each tool.
The doc subdirectory contains documentation and user’s guides for the tools.
The lib subdirectory contains linkable object libraries (*.a on UNIX systems, *.lib on MS DOS systems, and *.olb on VMS systems).
The source subdirectory contains source code and data dictionary files.
Note: the VMS version of the source code files are currently not available on-line. The source code files may be obtained by arrangement with the PDS Operator (see Appendix A for contact information)

3. Change directory to the subdirectory containing the desired files (e.g.“cd sunos/source”).
4. Change your ftp mode to binary (“binary”).
5. Get the files you need (“get<filename>”, or “mget *” to get all files in a subdirectory). Be sure to get the /toolbox_v4r1/readme file, which describes how to build each tool on your system.
6. Quit ftp (“quit”).
7. For UNIX systems, decompress and untar the files using the following commands:
   uncompress *.tar.Z
   tar -xvf *.tar
   For MS DOS systems, decompress the files using the following command:
   pkunzip *.*

8. For some of tools, the Data Dictionary files “pdsdd.full” and “pdsdd.idx” are needed. These are available in the source directories stand-alone and are also included with the source code. The Data Dictionary is updated whenever the on-line version changes (i.e., when new data elements or

objects are added or changes are made to existing elements or objects). VMS users (or anyone who has trouble using the Data Dictionary files after copying them) should run the make_index utility after copying the Data Dictionary file,”pdsdd.full”. The program “make_index” is among the programs in the sunos, dos, and vms subdirectories.

For VAX/VMS systems, running the make_index utility requires executing the following commands at the DCL prompt:

```
make_index:== “$[your directory]make_index.exe”
make_index pdsdd.full
```

For other systems, the command is simply:

```
make_index pdsdd.full
```

## B.3 PDS Label Library Light

The PDS Label Library Light (lablib3) is an evolving set of routines written in C which can be used to read, process, and write PDS labels. The Label Library reads and writes labels in the Object Description Language (ODL). The Label Library may be linked with any program which must deal with PDS labels. It was developed in order to encourage others to use PDS labels, to make it easier to mass produce PDS labels, and provide a method for enforcing ODL standards. The Label Library is a linking library which can be assimilated into any C program. It has been tested using the SUN C, VAX/VMS C, and Borland C++ compilers. Label Library Light is a reimplementation of Label Library Version 1.1 that is smaller and faster and that has a simpler function interface.

To obtain this library (which is currently available for SunOS only):

1. Anonymous ftp to starhawk.jpl.nasa.gov (137.79.108.113)

2. Change your ftp mode to binary (“binary”)

3. Get the file your need (“get lablib3.tar.Z”)

4. Quit ftp

5. Uncompress the file (“uncompress lablib3.tar.Z” )

6. Untar the resultant uncompressed file (“tar -xvf lablib3.tar”)

For additional information on the PDS Software Inventory, PDS Toolbox, or PDS Label Library Light, or for help accessing the files electronically, contact Steve Hughes at (818) 306-6030 or shughes@jplpds.jpl.nasa.gov (137.79.104.100), or contact the Planetary Data System Operator at pds_operator@jplpds.jpl.nasa.gov or JPLPDS::PDS_OPERATOR.

# Appendix C

# Glossary

**Attribute** (see Data Element)

**Browse Data Set**

A data set, typically of limited size and resolution, created to provide an understanding of the type and quality of data available in a related full resolution data set. It may also enable the quick selection of specific products or intervals for further processing or analysis. For an image data set, a browse data set could be a set of subsampled images of lower resolution. A browse data set for a time series sampled on 4 second intervals may consist of the data averaged at 48 second intervals. The form of browse data is generally unique for each type of data set and depends on the nature of the data and the criteria used for data selection within the related science discipline.

**Catalog Data**

Data that is obtained from data sets or data products, and which serves to describe, identify, associate, characterize, or highlight the data set. The values of the data elements in a catalog object are the catalog data, and are loaded into a PDS catalog to facilitate data searching and ordering.

**Catalog Object**

A set of required keywords with values used to provide data for a data product or data set that is suitable for loading into a PDS catalog. A catalog object template that is filled in with values becomes a catalog object. See Appendix B of the PDS Standard Reference for a complete set of PDS Catalog Object examples.

**Catalog Object Template**

An uncompleted catalog object; similar to a form. Once values are assigned to each of the data elements in the catalog object template, it becomes a catalog object.

**Data Element (Attribute, Element, Keyword)**

A data element is a characteristic of an object or group of objects provided in a PDS label. For example, in a label containing the line

TARGET_NAME = JUPITER

the data element is TARGET_NAME. Some data elements describe the physical or logical characteristics of the data (the number of ROWS in a table, for example). Others describe the content of the data (MINIMUM_WAVELENGTH of a spectrum, for example) or information to be used in interpreting the data (such as a UNIT of measure). The terms attribute, element, and keyword are often used interchangeably with data element.

**Data Object**

A data object is that portion of a data product that contains the actual data that is described in a data object definition within a PDS label. It is tangible, and can be physically accessed and manipulated. For example, if a PDS label contains the following lines,

```
^TABLE = “T142836.DAT”
...
OBJECT = TABLE
ROWS = 1028
...
END_OBJECT = TABLE
```

the data object for this table is in the file T142836.DAT. The actual data object is often classified as a primary data object or a secondary data object.

**Data Object Description**

A set of required and optional keywords which are used in a PDS label to define the structure and characteristics of a data object. Data object descriptions begin and end with the OBJECT = and END_OBJECT = lines. The PDS standard data object descriptions are defined in Appendix A of the PDS Standards Reference.

**Data Product**

A grouping of primary and secondary data objects and their associated PDS labels resulting from a scientific observation. A data product is a component of a data set.

**Data Product Label**

A set of data elements and values, data object definitions, and optional SFDUs, expressed in ODL that identifies, describes, and defines the structure and context of a data product.

**Data Set**

An accumulation of data products, secondary or ancillary data, software, and documentation, that completely document and support the use of those data products. A data set can be part of a data set collection.

**Data Set Collection**

A collection of data sets that are related by observation type, discipline, target, or time, and treated as a unit. Data set collections are archived and distributed together for a specific scientific objective.

**Element** (see Data Element)

**Keyword** (see Data Element)

**Label** (see Data Product Label)

**Logical Record**

A record independent of its physical environment, that exists from the standpoint of its content, function, and use rather than its physical attributes. It is defined in terms of the information it contains. Portions of the same logical record may be located in different physical records, or several logical records may be located in one physical record.

**Medium**

Any physical material capable of holding data (e.g., pages, film, magnetic tape, CD-ROM, etc.).

**Meta Data**

Selected or summary information about data. PDS catalog objects and data product labels are forms of meta data for summarizing important aspects of data sets and data products.

**Object**

An abstract or atomic entity that corresponds to something tangible in an archive environment, such as a grouping of data (data object) or a related grouping of information about that data(catalog object).

**Object Description Language** (ODL)

A specific language used to encode data product labels and catalog objects for the Planetary Data System and other NASA data systems. ODL was developed by PDS and is based upon the Flexible Image Transport (FITS) and Video Image Communication and Retrieval (VICAR) labelling systems used by both the astronomy and planetary image communities. (See the *Object Description Language Specification and Usage* chapter of the PDS Standards Reference).

**Physical Record**

A record whose characteristics depend on the manner or form in which it is stored, retrieved, or moved. A physical record may contain one or several logical records or a part of a logical record.

**Pointer**

A statement within a data product label or catalog object that identifies or references a starting point within the same file or an external file. Pointers are used to identify the start of a data object, the location of additional text for insertion, or the location of descriptive information for further reference.

**Primary Data Object**

The primary data resulting from a scientific observation in a data product. Examples of primary data objects are the actual data for a single IMAGE, or a TABLE containing one hour of data for an instrument. A secondary data object may be associated with the primary data object.

**Secondary Data Object**

A data object associated with a primary data object in a data product. A secondary data object may provide additional information or may be necessary for processing or correctly interpreting the primary data object. For example, an engineering table(secondary data object) is often associated with a time series (primary data object) of values from an instrument to provide important housekeeping information required to properly use the data.

**SFDU**

A SFDU (Standard Formatted Data Unit) is a data object that conforms to a specific set of CCSDS (the Consultative Committee for Space Data Systems) recommendations for structure, construction rules, and field specification definition. Details about the PDS usage of SFDUs can be found in the *SFDU Usage* chapter of the PDS Standards Reference.

**Standard Value**

A standard value is one of a set of possible values that can be assigned to a data element. These value sets are extendable for many data elements. Further information on the use and addition of standard values can be found in the Planetary Science Data Dictionary.

**Template** (see Catalog Object Template)

**Volume**

A volume is a unit of physical media such as a CD-ROM or CD-WO, an 8mm magnetic tape, or a 9-track magnetic tape. Data sets may reside on one or more volumes and multiple data sets may also be stored on a single volume. Volumes may be grouped into volume sets.

**Volume Set**

A grouping of one or more related volumes treated as a unit.

# Appendix D

# Acronyms

The following list contains acronyms and abbreviations used throughout PDS documentation.

| | |
|---|---|
| AMMOS | Advanced Multi-Mission Operations System |
| ANSI | American National Standards Institute |
| APDTP | Archive Policy and Data Transfer Plan |
| ARC | Ames Research Center |
| ASCII | American Standard Code for Information Interchange |
| ASU | Arizona State University |
| CCSDS | Consultative Committee for Space Data Systems |
| CD-ROM | Compact Disc - Read Only Memory |
| CD-WO | Compact Disc - Write Once |
| CIT | California Institute of Technology |
| CN | PDS Central Node |
| CODE SL | NASA’s Solar System Exploration Division |
| CODMAC | Committee on Data Management and Computation |
| COSMIC | Computer Software Management & Information Center |
| CR | Carriage Return |
| DAT | Digital/Audio Tape |
| DBA | Database Administrator |
| DBMS | Database Management System |
| DDL | Data Distribution Laboratory |
| DEC | Digital Equipment Corporation |
| DECNET | DEC Network |
| DIF | Directory Interchange Format |
| DIM | Digital Image Model |
| DN | PDS Discipline Node |
| DOS | Disk Operating System |
| DPW | Data Preparation Workbook |
| DTM | Digital Terrain Model |
| EDR | Experiment Data Record |
| Email | Electronic Mail |
| EPS | Encapsulated Postscript |
| FILE2LAB | PDS File to Label Generator |
| FITS | Flexible Image Transport System |
| FTP | File Transfer Protocol |
| GDS | Ground Data System |
| GMT | Greenwich Mean Time |
| GSFC | Goddard Space Flight Center |
| GTDA | General Technical Data Available |

| | |
|---|---|
| IAU | International Astronomical Union |
| IBM | International Business Machines |
| ID | Identifier |
| IDS | Interdisciplinary Scientist |
| IEEE | Institute of Electrical and Electronics Engineers |
| IR | Infrared Radiometer |
| IRPS | Image Retrieval and Processing System |
| IRTM | Infrared Thermal Mapper |
| ISO | International Standards Organization |
| ISIS | Integrated Software for Imaging Spectrometers |
| JD | Julian Date |
| JPL | Jet Propulsion Laboratory |
| LASP | Laboratory for Atmospheric and Space Physics |
| LF | Line Feed |
| LSB | Least Significant Byte first |
| MAC | Macintosh |
| MDR | Master Data Record |
| MD | Master Directory |
| MGDS | Multimission Ground Data System |
| MIF | Maker Interchange Format |
| MIFT | Mission Interface Team |
| MIPS | Multimission Image Processing Subsystem |
| MIT | Massachusetts Institute of Technology |
| MOU | Memorandum of Understanding |
| MSB | Most Significant Byte first |
| MS-DOS | Microsoft Disk Operating System |
| N/A | Not Applicable |
| NAIF | Navigation and Ancillary Information Facility |
| NASA | National Aeronautics and Space Administration |
| NBS | National Bureau of Standards |
| NRC | NASA Research Counci |
| NSI | NASA Science Internet |
| NSSDC | National Space Science Data Center |
| ODL | Object Description Language |
| ODR | Original Data Record |
| OSSA | Office of Space Science and Applications |
| PC | Personal computer |
| PCWG | Planetary Cartography Working Group |
| PDB | Project Data Base |
| PDMP | Project Data Management Plan |
| PDS | Planetary Data System |
| PDSMC | PDS Management Council |
| PI | Principal Investigator |
| PICS | Planetary Image Cartography System |
| PPI | Planetary Plasma Interactions Node |
| PSDD | Planetary Science Data Dictionary |

| | |
|---|---|
| PS | Postscript |
| PVL | Parameter Value Language |
| RADIG | Radar Investigation Group |
| RDBMS | Relational Database Management System |
| RDR | Reduced Data Record |
| RPIF | Regional Planning Image Facility |
| RTOP | Research and Technology Objectives and Plans |
| SBN | Small Bodies Node |
| SEDR | Supplementary Experiment Data Record |
| SFDU | Standard Formatted Data Unit |
| SGML | Standard Generalized Markup Language |
| SI | System Internationale d'Unites |
| SIS | Software Interface Specification |
| SPAN | Space Physics Analysis Network |
| SPICE | Spacecraft, Planetary ephemeris, Instrument pointing, C-matrix, Events |
| SQL | Structured Query Language |
| TAB2LAB | PDS Table to Label Generator |
| TCP/IP | Transmission Control Protocol/Internet Protocol |
| TeX | Automated Documention Program (Tau Epsilon Xhi) |
| TIFF | Tagged Image File Format |
| UCLA | University of California, Los Angeles |
| UNK | Unknown |
| USGS | United States Geological Survey |
| UTC | Universal Time Coordinated (often called GMT) |
| UV | Ultraviolet |
| VAX | Virtual Address/Access Extension (DEC computers) |
| VICAR | Video Communication and Retrieval System |
| VMS | Virtual Memory System |
| WAIS | Wide Area Information Service |
| WORM | Write Once Read Many |
| WWW | World Wide Web |